

Extending ENSO Prediction beyond 18 Months through Stage-consistent Ocean Internal Dynamics Based on Deep Learning

Yu Weng

School of Atmospheric Sciences, Nanjing University of Information Science and Technology

Xiefei Zhi

`zhi@nuist.edu.cn`

School of Atmospheric Sciences, Nanjing University of Information Science and Technology

Jingyu Wang

`jingyu.wang@nie.edu.sg`

National Institute of Education, Nanyang Technological University

Ling Zhang

School of Atmospheric Sciences, Nanjing University of Information Science and Technology

Yang Lyu

School of Atmospheric Science and Remote Sensing, Wuxi University

Yan Ji

School of Atmospheric Science and Remote Sensing, Wuxi University

Yunchang Cao

School of Atmospheric Sciences, Nanjing University of Information Science and Technology

Shoupeng Zhu

Nanjing Innovation Institute for Atmospheric Sciences, Chinese Academy of Meteorological Sciences–
Jiangsu Meteorological Service

Klaus Fraedrich

Max Planck Institute for Meteorology

Dexuan Kong

Guizhou Institute of Mountainous Meteorological Sciences

Liqun Zhou

School of Atmospheric Sciences, Nanjing University of Information Science and Technology

Research Article

Keywords: sea subsurface temperature, ENSO, CNN, ConvTransformer, deep learning

Posted Date: March 31st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9268042/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Extending ENSO Prediction beyond 18 Months through Stage-consistent Oceanic Internal Dynamics Based on Deep Learning

Yu Weng^{1,2}, Xiefei Zhi^{*1,2}, Jingyu Wang^{†3}, Ling Zhang^{1,2,4}, Yang Lyu⁵, Yan Ji⁵,
Yunchang Cao^{1,2}, Shoupeng Zhu^{6,7}, Klaus Fraedrich⁸, Dexuan Kong⁹, Liqun Zhou^{1,2}

¹ State Key Laboratory of Climate System Prediction and Risk Management/Key Laboratory of Meteorological Disaster, Ministry of Education/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044, China

² School of Atmospheric Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China

³ National Institute of Education, Nanyang Technological University, Singapore 637616, Singapore

⁴ Information Center, Ministry of Water Resources of China, Beijing 100053, China

⁵ School of Atmospheric Science and Remote Sensing, Wuxi University, Wuxi 214105, China

⁶ Nanjing Innovation Institute for Atmospheric Sciences, Chinese Academy of Meteorological Sciences – Jiangsu Meteorological Service, Nanjing 210044, China

⁷ Key Laboratory of Transportation Meteorology of CMA / Jiangsu Key Laboratory of Severe Storm Disaster Risk, Nanjing 210044, China

⁸ Max Planck Institute for Meteorology, Hamburg 20146, Germany

⁹ Guizhou Institute of Mountainous Meteorological Sciences, Guiyang 550002, China

*Corresponding author. Email: zhi@nuist.edu.cn

†Corresponding author. Email: jingyu.wang@nie.edu.sg

Abstract: Deep learning (DL) has emerged as a transformative tool in atmospheric science, demonstrating particular efficacy in El Niño-Southern Oscillation (ENSO) prediction by achieving accurate forecasts with a notable lead time. Nevertheless, the adoption of DL-based ENSO forecasting models remains constrained by insufficient physical interpretability. In this study, we leveraged the linkage between sea subsurface temperature (SUBT) propagation and ENSO development — namely, a propagation loop in equatorial/extra-equatorial Pacific subsurface temperatures where anomaly signals propagate over timescales matching ENSO interannual recurrence—to develop two physically constrained, lightweight, and generalizable DL models for ENSO prediction: (i) a Convolutional Neural Network (CNN) architecture integrated with a Convolutional Block Attention Module (CBAM), and (ii) a ConvTransformer model built upon this CNN framework. These models effectively reconstruct SUBT anomaly propagation and skillfully forecast the Niño3.4 index at 18-month lead times. Guided by the physical interpretation of the ConvTransformer model using Integrated Gradients, we proposed a novel ENSO stage partitioning scheme that is physically consistent with the propagation of SUBT anomaly, providing a robust explanation for the onset, evolution, and transition of El Niño and La Niña events within a unified dynamical framework. This DL-derived scheme not only enhances forecast reliability but also offers new insights into the subsurface-driven mechanisms underlying ENSO variability and predictability.

Keywords: sea subsurface temperature, ENSO, CNN, ConvTransformer, deep learning

1. Introduction

The El Niño-Southern Oscillation (ENSO), the most prominent interannual variability signal in the tropical Pacific Ocean, is manifested through oscillations in sea surface temperature (SST) and sea level pressure in the equatorial Pacific¹. It has attracted widespread attention across oceanographic, meteorological, and environmental sciences due to its significant impact on global climate and ecological systems²⁻⁶. ENSO can affect the mid-latitude atmospheric circulation through atmospheric remote correlation (e.g., Pacific–North American Pattern [PNA], North Atlantic Oscillation [NAO]), which in turn causes large-scale global climate anomalies^{7,8}. Furthermore, within the framework of global warming, the capacity of ENSO to instigate atmospheric anomalies in the extra-tropics has been augmented^{9,10}, while a shallower thermocline and stronger upper-ocean stratification in the tropical eastern Pacific now drive

59 more frequent multi-year ENSO events and prolonged cycles^{11,12}. Given the pivotal role of
60 ENSO in modulating the global climate, the advancement of both ENSO dynamical and
61 statistical models, as well as the enhancement of ENSO prediction proficiency, are of utmost
62 necessity.

63 Traditionally, ENSO prediction have been produced using either statistical models with
64 restrictive assumptions, such as a linear relationship between predictors and predictands¹³, or
65 computationally intensive physics-based dynamic models^{14,15}, or a combination of both^{16,17}.
66 Over the past four decades, many statistical or dynamical models have been developed for real-
67 time ENSO prediction. Zebiak and Cane¹⁸ developed the first successful dynamic model for
68 seasonal ENSO prediction, laying the foundation for ENSO research. Subsequent studies have
69 continuously improved ENSO predictive capabilities. Statistical or dynamic models use gridded
70 sea surface temperature, wind and pressure fields, as well as sea subsurface temperature (SUBT)
71 data, and can forecast ENSO 6 to 12 months in advance¹⁹⁻²⁶. However, neither statistical nor
72 dynamical models can generate reliable forecasts of ENSO prior to the boreal summer—a
73 limitation known as the Spring Predictability Barrier (SPB)²⁷.

74 Over the past two decades, the rapid advancement of artificial intelligence (AI) and deep
75 learning (DL) has extended ENSO prediction to longer lead times. Ham et al.²⁸ first
76 demonstrated 18-month lead ENSO prediction feasibility using a three-layer Convolutional
77 Neural Network (CNN) model, later advanced by Ye et al.²⁹ with a novel parallel deep CNN
78 featuring heterogeneous architecture. This design adapts network structures to different forecast
79 lead times, enabling variability modeling of distinct lead-time tasks, thereby enhancing the
80 reliability of long-term ENSO prediction. Gupta et al.³⁰ expanded the utility of ConvLSTM
81 models, skillfully predicting monthly mean Niño3.4 index with a 12-month lead time, while
82 Wang et al.³¹ proposed an integrated framework combining a ConvLSTM model with an
83 improved attention mechanism and genetic algorithm (GA)-based model recombination to
84 enhance ENSO prediction capabilities in CMIP5/6 models. Recent research has explored
85 synergies between CNNs and transformers in ENSO prediction. For example, Lyu et al.³²
86 reported that ResoNet—a deep learning model leveraging hybrid convolution and transformer
87 networks—achieved an average forecast lead time of 18 months for the Niño3.4 index, with
88 ideal-case predictions extending to 26 months. These findings collectively highlight CNNs as a

versatile toolset for ENSO prediction. Notably, while DL models have demonstrated remarkable prowess, dynamical frameworks remain indispensable in climate forecasting. Chen et al.³³ addressed this duality with a hybrid dynamical-DL model, significantly improving ENSO forecast skill across all seasons and lead times.

Built on these advances, DL-based ENSO modeling is evolving along three axes: (1) increasing architectural complexity, (2) multi-domain data integration, and (3) expanded spatial domains—from the tropical Pacific to remote regions linked via teleconnections³⁴⁻³⁷. There are two major drawbacks of such aggressive expansion. The first being the potential conflation of causal drivers with resultant signals. Many such signals are not independent precursors of ENSO but are instead consequences or remote responses to it. Incorporating these ENSO - induced signals as predictors creates a risk of data leakage, where the model implicitly "learns" from information it would not have access to in a true forecasting context. This undermines the model's causal integrity and predictive robustness, as it may perform well in retrospective validation but fail in true forecasting scenarios where future response signals are not yet realized. Moreover, this practice can blur the physical interpretability of the model, making it difficult to distinguish genuine precursors from feedbacks or noise³⁸. The second drawback is that the advantage of DL is diminishing³⁹—it is no longer an economic and fast substitute for dynamical models. As the DL architectures become increasingly complex, with high-dimensional inputs, large training datasets, and longer lead times, the computational cost of training and tuning these models begins to rival that of traditional dynamical systems. Moreover, without strong physical constraints⁴⁰, these models often require extensive trial-and-error experimentation, hyperparameter optimization, and retraining to achieve robust performance across different ENSO phases. In some cases, the interpretability and physical consistency offered by dynamical models may outweigh the marginal skill gains of deep learning models, especially when their complexity leads to issues such as overfitting, instability, or poor generalizability across climate regimes.

Given these limitations, there exists an opportunity to renew emphasis on physically interpretable oceanic mechanisms—not as a replacement for data-driven approaches, but as a complementary framework that could provide comparable predictive skill but enhance physical clarity. Among the widely recognized theoretical frameworks explaining ENSO's underlying

mechanisms in the scientific community are the Bjerknes positive feedback mechanism⁴¹, the trade wind relaxation theory⁴², the delayed oscillator theory⁴³, and the recharge oscillator theory^{44,45}. Collectively, these foundational studies delineate the core dynamical processes governing ENSO evolution: Anomalous ocean temperature in the western equatorial Pacific propagate eastward under the combined influence of anomalous trade winds and oceanic currents—particularly the Equatorial Undercurrent, which facilitates the zonal advection of thermal anomalies toward the eastern Pacific. Concurrently, atmospheric forcing over the extratropical central Pacific excites westward-propagating baroclinic Rossby waves; these waves, carrying subsurface thermal and dynamical anomalies, traverse the Pacific basin over a period of 1-2 years before reaching the western Pacific coast, where they modulate the thermocline depth and subsequently feed back onto the equatorial SST anomalies, sustaining the ENSO cycle. Complementing these overarching theoretical frameworks, a large body of subsequent studies has elaborated on the robust link between ENSO variability and oceanic internal dynamics. Early ENSO theory largely framed tropical climate variability as an internal tropical air–sea interaction^{41,46}—e.g., zonal warm-water redistribution—while influential intermediate models typically held the equatorial thermocline structure fixed (mean depth and vertical stratification)^{18,47}, thereby underrepresenting sensitivity to high-latitude water-mass exchange. Similarly, although the delayed oscillator theory⁴³ identifies the delayed feedback effect of extra-equatorial Rossby wave propagation on equatorial SST, it does not precisely define the specific range of the extra-equatorial Pacific, further limiting the mechanistic clarity of cross-equatorial interactions in ENSO dynamics. Addressing this theoretical gap, Gu and Philander⁴⁸ proposed a core argument that cross-latitudinal exchanges between the tropics and extratropics serve as the primary driver of interdecadal fluctuations, highlighting the need to incorporate extratropical processes into the understanding of long-term climate variability. Meanwhile, Lysne et al.⁴⁹ focused specifically on the mechanisms and pathways through which mid-to-high latitude subsurface temperature anomalies signals in the North Pacific influence climate variability in the equatorial region. Their findings demonstrated the existence of rapid subsurface ocean connectivity between the subtropics and tropics: temperature anomalies assimilated in the subtropical zone can propagate to the equatorial region on an interdecadal timescale, ultimately modulating equatorial variability by altering key thermal properties of the

equatorial thermocline. This finding underscored the North Pacific's dominant role in modulating ENSO dynamics. Qian and Hu⁵⁰ further refined these insights by documenting distinct pathways of sea subsurface temperature anomalies: eastward propagation along 8°S–8°N and westward propagation along 8°N–24°N. Notably, the periodicity of signals traversing this loop aligns with the interannual recurrence of El Niño events, establishing a robust physical link between subsurface thermal dynamics and ENSO timing. Additionally, experiments demonstrate that Pacific subsurface temperature serves as a carrier of long-lead ENSO signals^{51–57}. Collectively, these studies highlight the predictive value of sea subsurface temperature in both equatorial and extratropical Pacific regions, positioning them as viable precursors for ENSO prediction. Notably, recent advances in DL-based ENSO prediction have begun to leverage subsurface-related variables: for instance, Zhou and Zhang⁵⁸ integrated sea surface wind stress and upper-ocean temperature anomaly fields into their self-attention-based neural network to achieve skillful ENSO prediction, while Zhao et al.³⁸ demonstrated that incorporating warm water volume (WWV)—a key integrated metric that quantifies the cumulative magnitude of subsurface thermal anomalies across the tropical Pacific—significantly enhanced ENSO prediction skill. However, as an integrated single-value metric, WWV inherently lacks the capacity to capture the spatiotemporal propagation characteristics of subsurface ocean temperature anomalies (e.g., zonal propagation along the equator or meridional transmission between the equatorial and extratropical regions), which are critical for understanding the dynamical processes driving ENSO evolution. Furthermore, these aforementioned studies have explicitly accounted for the nonlinear air-sea interactions; in contrast, this study focuses exclusively on exploring the synergistic predictive potential of coordinated subsurface temperature variability across the equatorial and extratropical Pacific within current DL modeling frameworks. Building on these insights, this study develops two fundamentally simple DL models using only SST and SUBT as inputs to probe the source of ENSO predictability. Contrary to conventional expectations, our results reveal that: (i) forecast lead times can be substantially extended, and (ii) all predictive improvements originate exclusively from SUBT. More importantly, by distilling knowledge from these DL models, we derive a novel ENSO stage partitioning scheme that demonstrates robust dynamical linkages to SUBT variability. These findings not only challenge prevailing assumptions about ENSO

prediction but also pave the way for developing next-generation, physics-informed simplified ENSO models that leverage subsurface thermal structure as a key ENSO predictor.

2. Results

2.1. Improved ENSO deep learning forecast driven by subsurface dynamics

Building on the relationship between sea subsurface temperature anomalies propagation and ENSO evolution, we assessed the ENSO forecasting skill of the ConvTransformer model across three distinct input configurations to investigate the importance of sea subsurface temperature in ENSO prediction: (i) sea surface temperature (SST) anomalies alone, (ii) sea subsurface temperature (SUBT) anomalies alone, and (iii) combined SST anomalies and SUBT anomalies. Additionally, for the CNN model, we used both SST and SUBT anomalies as input data to explore differences in forecasting skill between the two deep learning models. The analysis of the 18-month forecast lead time, as measured by anomaly correlation coefficient (ACC), root mean square error (RMSE), conditional RMSE, and weighted RMSE, reveals several key findings:

Figure 1a shows that the SST-only configuration demonstrates limited forecast capability, with correlation coefficients exceeding 0.5 only for lead times up to 8 months, indicating poor long-term predictability. In contrast, the ConvTransformer model shows significantly enhanced skill when utilizing either sea subsurface temperature anomalies alone or combined SST and sea subsurface temperature data. Most notably, the hybrid SST-subsurface configuration maintains correlation coefficients above 0.5 throughout the entire 18-month forecast lead time, with these correlations being statistically significant, representing a substantial improvement over surface-only predictions. Notable is also the combined SST and sea subsurface temperature anomaly configuration, which yields superior performance attaining a remarkable near-term skill ($r=0.96$ at 1-month lead time). A comparison of the three types of root mean square error (RMSE) reveals that, for short lead times, the differences in these three RMSE metrics across the three input configurations are minimal; however, for longer lead times, the configurations incorporating sea subsurface temperature yield better forecasting performance, with all three RMSE values lower than those of the ConvTransformer model using only SST. These substantial improvements highlight the critical role of sea subsurface temperature in ENSO evolution, as the additional thermocline information provides crucial memory effects and

precursor signals that SST data alone cannot capture.

A further comparison of the performance between the ConvTransformer and the CNN models (utilizing both SSTA and sea subsurface temperature anomalies), the seven individual North American Multi-Model Ensemble (NMME) models, and the equal-weighted ensemble of the seven NMME models (hereafter referred to as MME) (Fig. 1a) reveals that, relative to traditional climate models, the ConvTransformer and CNN models exhibit superior correlation-based skill at lead times of 6 months and 8 months, respectively. Notably, for long-term predictions with lead times exceeding 9 months, both deep learning models outperform the MME significantly. Additionally, while the CNN model initially shows lower correlation coefficients than the ConvTransformer model, it ultimately surpasses the latter's skill at longer lead times. These results collectively highlight that deep learning models can improve both the lead time and accuracy of ENSO prediction compared to numerical climate models.

The comparison of full-sample RMSE (Fig. 1b) reveals that, relative to climate models, deep learning models driven by both sea surface temperature anomalies and sea subsurface temperature anomalies exhibit larger RMSE values in the early forecast lead times, with these differences being statistically significant compared to MME. However, for lead times of 10–12 months, their RMSE is lower than that of all climate models and the MME. Notably, for lead times of 9–12 months, the CNN model's RMSE is significantly lower than that of MME. Additionally, across all forecast lead times, the RMSE of the CNN model is lower than that of the ConvTransformer model.

For ENSO research, strong ENSO events carry greater scientific significance, and an exclusive reliance on full-sample RMSE offers only limited evaluation value. To specifically assess model performance in forecasting strong ENSO events (defined as $|\text{Niño3.4}| > 1.0^{\circ}\text{C}$), conditional RMSE was employed (Fig. 1c). The results indicate that for both climate models and the CNN model, conditional RMSE during extreme phases is higher than full-sample RMSE, with RMSE increasing more rapidly as forecast lead time extends—reflecting larger prediction errors and greater challenges in forecasting during strong ENSO events. This phenomenon is consistent with previous studies highlighting more complex air-sea coupling and nonlinear feedback mechanisms during extreme ENSO events⁵⁹. Notably, in the early forecast lead times, the conditional RMSE of two ConvTransformer configurations—one driven solely by sea

subsurface temperature anomalies and the other by both SSTA and sea subsurface temperature anomalies—is significantly lower than their full-sample RMSE. By the 3-month and 5-month lead times, the conditional RMSE of these two configurations even lower than that of MME and other climate models, respectively; furthermore, the differences in conditional RMSE relative to the MME are statistically significant. Furthermore, when focusing exclusively on strong ENSO events, the conditional RMSE of these two ConvTransformer configurations is markedly lower than that of the CNN model, indicating that the ConvTransformer model achieves more accurate predictions for extreme events, whereas the CNN model performs more stably during neutral ENSO phases.

To emphasize the importance of extreme phases in full-sample verification, we further introduced weighted RMSE (Fig. 1d). It is found that, for forecast lead times of 9-12 months, the weighted RMSE of both the ConvTransformer model and the CNN model driven by both SST anomalies and sea subsurface temperature anomalies is significantly lower than that of MME. Additionally, across the entire range of forecast lead times, the weighted RMSE of the ConvTransformer model is consistently lower than that of the CNN model. Combined with the sensitivity analysis in Fig. S15, it can be further seen that, except for the ConvTransformer model, the weighted RMSE of all models increases as the weight assigned to strong events becomes larger, and this increase is more pronounced at medium and long lead times. For the CNN model in particular, weighted RMSE increases systematically with increasing weight, with a more marked rise beyond 12 months. In contrast, the ConvTransformer model shows relatively weak sensitivity to weight changes and retains the lowest or nearly the lowest weighted RMSE at most lead times. Meanwhile, persistence and the MME are also sensitive to changes in weighting, with their error curves increasing more rapidly at medium and long lead times. These results indicate that deep learning models incorporating subsurface information are generally superior to conventional reference forecasts and that, within an evaluation framework that emphasizes the contribution of strong events, the advantage of the ConvTransformer model over the CNN model becomes more evident. Taken together, the results from Figs. 1 and 2-5 indicate that, without relying on extremely large training datasets or complex network extensions, the models developed in this study achieve statistically skillful ENSO prediction out to 18 months. This demonstrates that subsurface temperature information

is a key predictor for improving medium- and long-range ENSO forecasting skill.

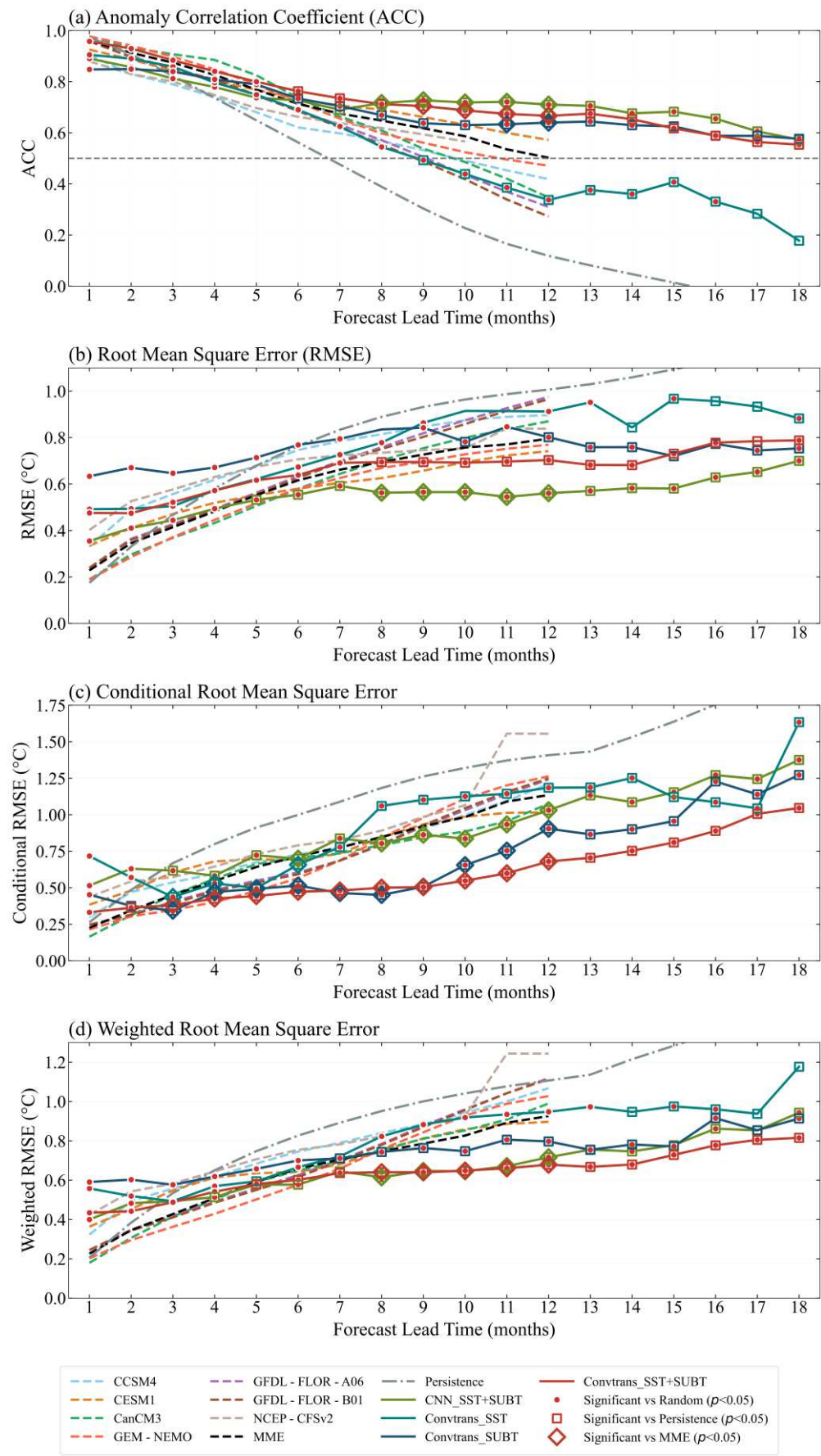


Fig. 1 Performance of ConvTransformer, CNN and other climate models: (a) all-season ACCs of ConvTransformer, CNN, seven NMME models, and the equal-weighted ensemble of the seven NMME models (MME) at different forecast lead times; (b) as in (a) but using RMSE as the performance metric; (c) conditional RMSE of ConvTransformer, CNN, seven NMME models, and the equal-weighted ensemble of the seven NMME models (MME) at different forecast lead times, where extreme events are defined using a fixed threshold τ (with $\tau = 1.0^{\circ}\text{C}$ for the Niño3.4 index); (d) the weighted RMSE of ConvTransformer, CNN, seven NMME models, and the equal-weighted ensemble of the seven NMME models (MME) at different forecast lead times, where extreme events are defined using a fixed threshold τ (with $\tau = 1.0^{\circ}\text{C}$ for the Niño3.4 index) and a weight of 3.0 is assigned to extreme events while a weight of 1.0 is assigned to non-extreme events. The bootstrapping method was employed to test the significance of the ACC and the three types of RMSE derived from the forecast outputs of the deep learning models (ConvTransformer and CNN). Solid red dots indicate results that are significantly better than the random forecast benchmark at the 95% confidence level, open squares indicate results that are significantly better than the persistence benchmark at the 95% confidence level, and open diamonds indicate results that are significantly better than the MME at the 95% confidence level.

The forecast performance is better illustrated in Figure 2, which compares the predicted and observed values from both ConvTransformer model and CNN model using combined SST anomalies and SUBT anomalies. Both deep learning models successfully capture the monthly variation of the actual Niño3.4 index, though systematic biases emerge in magnitude estimation, which is particularly evident in the underprediction of the 2015 El Niño event intensity. In addition, forecast skill exhibits pronounced lead-time dependence, with initial phase deviations appearing at 6-month lead time that amplify into complete phase reversals as lead time extends.

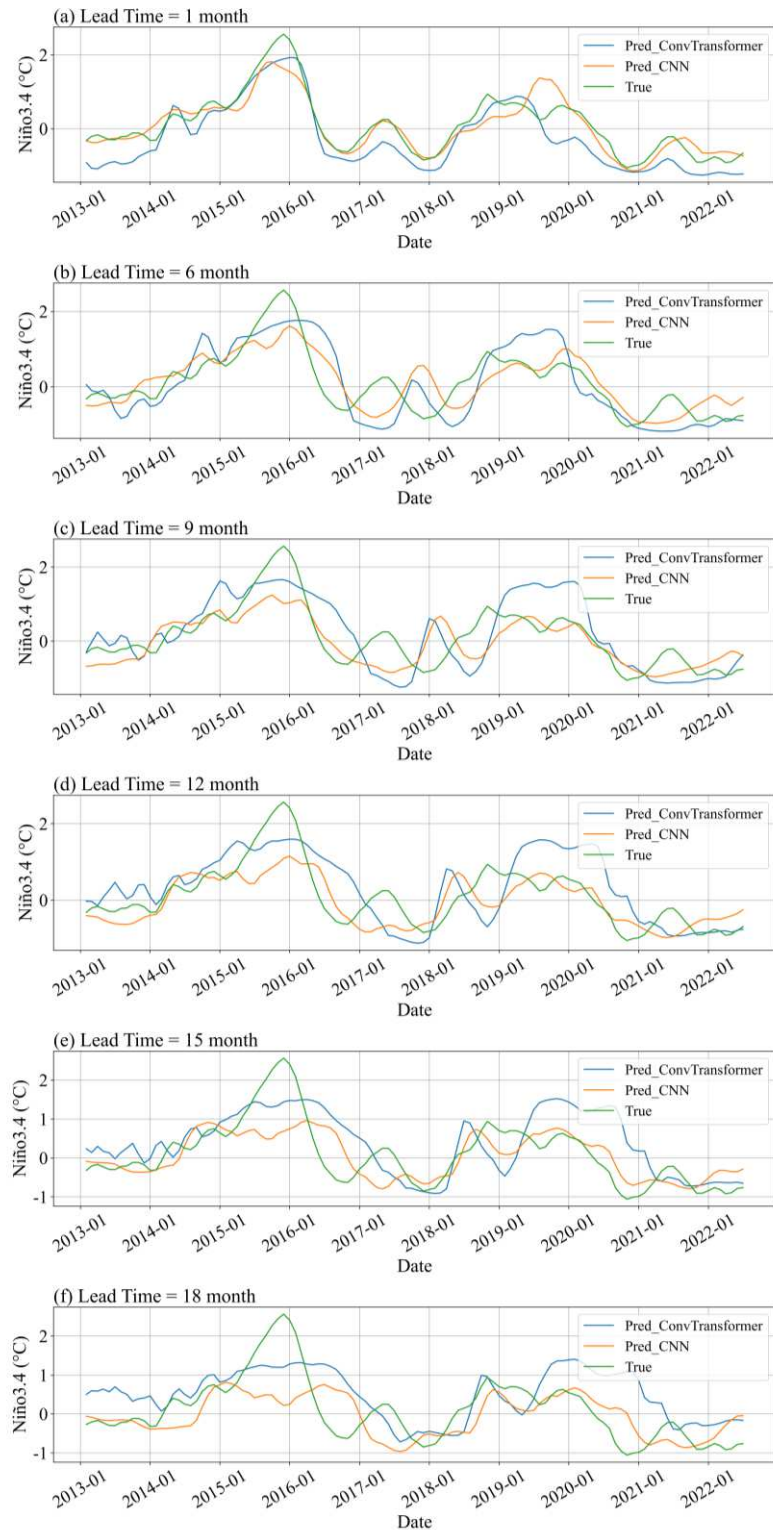


Fig. 2 Monthly time series (2013-02 to 2022-07) of the actual and predicted Niño3.4 index at 1-, 6-, 9-, 12-, 15-, and 18-month lead times. Above-average (below-average) values indicate El Niño (La Niña) events, respectively.

Comparative analysis of two strong El Niño events (2015, 2023) (Fig. S1) and two La Niña events (2017, 2021) (Fig. S2) reveals distinct performance characteristics between the two deep learning models across ENSO events. As illustrated in Figure S1, both the ConvTransformer and CNN models effectively capture the evolution of strong El Niño events, with the ConvTransformer model outperforming the CNN—its predictions remain above the threshold of 1.0 °C even at 18-month lead time, whereas the CNN model exhibits a more significant attenuation in predictive amplitude as lead times extend—though its predictions remain valid overall. In contrast, for La Niña predictions (Fig. S2), both models exhibit significantly larger temporal lags, with phase delays accumulating to several months at extended lead times.

This study further assessed the forecasting skill of two deep learning models and MME for El Niño and La Niña events using three key metrics: Hit Rate (HR), False Alarm Rate (FAR), and Critical Success Index (CSI). To address uncertainties arising from the small sample size of ENSO events, Table S1 presents the bootstrapped mean HR, FAR, and CSI values for the ConvTransformer model, CNN model, and MME, while Table S2 shows the corresponding confidence intervals based on interquartile ranges (IQR). To verify the significance of differences in HR between the deep learning models and MME, the bootstrapping method was employed to conduct a significance test on their HR differences (Δ HR) (Fig. S3). The evaluation results in Table S1 indicate that, except for the 1-month lead time forecast of El Niño events, the HR values of the two deep learning models are significantly higher than that of MME at lead times of 6-12 months (Fig. S3), indicating a substantial and reliable improvement in predictive skill. For La Niña event prediction, the HR of the ConvTransformer model is significantly higher than that of MME at lead times of 1, 6, 9, and 12 months (Fig. S3), while the HR of the CNN model is also significantly higher than that of MME at lead times of 1 and 12 months (Fig. S3). Additionally, within the 18-month forecast lead time, both the ConvTransformer and CNN models demonstrate excellent forecasting skill for El Niño events (i.e., high HR, low FAR, and large CSI). Specifically, at the 18-month lead time, the HR values of the two models remain as high as 85% (ConvTransformer) and 74% (CNN). However, forecasting La Niña events poses greater challenges: as the forecast lead time extends, the HR exhibits a more pronounced downward trend. Specifically, the HR of the ConvTransformer model decreases from 100% to 67%, while that of the CNN model drops from 79% to 65%. To

compare the significance of HR differences between the two deep learning models in forecasting El Niño and La Niña events, Fig. S4 presents the 95% confidence intervals (CIs) of Δ HR for the two deep learning models in forecasting these two events. Open pentagrams indicate that the differences passed the 95% significance test; it can be observed that the HR of the ConvTransformer model is significantly higher than that of the CNN model at lead times of 6, 9, 12, and 18 months.

Additionally, Table S1 shows that the two deep learning models maintain relatively stable forecasting capability for El Niño events across all forecast lead times, with their performance for El Niño being superior to that for La Niña events. To assess the significance of differences in HR between the models' predictions of El Niño and La Niña events, Fig. S5 presents the 95% confidence intervals (CIs) of Δ HR for the models' predictions of these two events, where open pentagrams indicate that the differences passed the 95% significance test. It can be observed that, at forecast lead times of 6, 9, 12, and 18 months, the HR values of both deep learning models for El Niño event predictions are significantly higher than those for La Niña event predictions. This phenomenon is consistent with conclusions from previous observational studies, which have documented differences in predictability between the warm ENSO phase (El Niño) and the cold ENSO phase (La Niña)⁶⁰. As demonstrated by Hayashi and Jin⁶¹, subsurface cold anomalies originating in the western Pacific weaken significantly by the time they propagate to the easternmost Pacific; this attenuation reduces the magnitude of surface cooling in the eastern Pacific, thereby dampening the intensity of subsequent La Niña events. Based on this observational and analytical evidence, their study concluded that the subsurface nonlinear dynamical heating contributes to asymmetry in both the magnitude and structure between El Niño and La Niña. Therefore, the observed differences in our models' predictive performance for El Niño versus La Niña events may stem from two key factors: first, the stronger coherent coupling between the equatorial and subtropical subsurface ocean during El Niño events, which enhances the traceability of predictive signals; and second, the presence of phase biases in the model predictions. These phase biases are likely attributed to two compounding effects: the larger number of historical samples available for warm events compared to cold ones, and the lower signal-to-noise ratio of La Niña signals⁶².

2.2. Spring Predictability Barrier (SPB)

To investigate the overall seasonal dependence of forecast values, this study calculated the monthly correlations of the Niño3.4 index predicted by the ConvTransformer and CNN models—with inputs of sea surface temperature anomalies and sea subsurface temperature anomalies—across forecast lead times of 1, 6, 9, 12, 15, and 18 months (Fig. S6). To mitigate the effects of limited sample size and to provide a robust estimation of the statistical significance of the monthly correlations, the bootstrapping method was employed to conduct statistical significance testing on the monthly correlation values, and all correlations were found to be statistically significant at the 95% confidence level.

From the monthly variations in the ACC, it is evident that the ACC values of both deep learning models consistently decline between April and May across all forecast lead times—a pattern consistent with the typical manifestation of the Spring Predictability Barrier (SPB). This aligns with the finding of Ham et al.²⁸, who noted that model forecasts tend to perform better during the early months of the calendar year. However, critical differences in the magnitude of this April–May ACC decline, combined with comparisons against traditional methods, highlight the superiority of the artificial intelligence (AI)-based approach developed in this study: for short-to-medium lead times (1-12 months), the April-May ACC reduction is extremely modest, indicating minimal degradation of predictive skill during spring; furthermore, when benchmarked against conventional ENSO forecasting methods⁶³⁻⁶⁵, both deep learning models achieve substantially higher spring predictive skill—specifically, at the 6-month lead time, the ConvTransformer and CNN models yield ACC values of 0.75 and 0.78, respectively, in May, while at the 12-month lead time, their ACC values remain as high as 0.69 and 0.60. These values outperform the spring ACC levels of traditional methods reported in the literature. Even for longer lead times (15-18 months), where the April-May ACC decline becomes more pronounced, the models maintain robust predictive skill in May, with ACC values of 0.68 (ConvTransformer) and 0.64 (CNN) at the 15-month lead time. It is evident that the long-lead predictions of the two deep learning models also outperform the correlation coefficients achieved by several other deep learning models during the period from April to May^{28,29}. Collectively, these results demonstrate that the deep learning models—incorporating sea subsurface temperature data for ENSO prediction—can effectively mitigate the SPB. The ability of the models to alleviate this barrier stems from the physical information embedded in

sea subsurface temperature data: unlike SST signals, which are highly susceptible to interference from seasonal atmospheric variability (a key driver of the SPB)⁶⁶, sea subsurface temperature exhibits coherent propagation characteristics—its signals can be stably transmitted across seasons (from winter to spring, and further to summer and autumn) without significant attenuation. This enables the deep learning models to continuously capture and learn ENSO-related key physical features throughout the spring season, circumventing the "signal loss" issue that plagues traditional models and thereby sustaining stable predictive skill during spring.

Figures. S7 and S8 present the seasonal growth rates of forecast errors $\bar{k}$ for the ConvTransformer and CNN models, respectively, during the growing and decaying periods of the 2015/2016 El Niño event and 2017/2018 La Niña event across different forecast lead times. For the ConvTransformer model's predictions of the El Niño growing period (Fig. S7a), $\bar{k}_{AMJ}$ is mostly negative, while $\bar{k}_{JAS}$ is positive; $\bar{k}_{max}$ occurs predominantly in the JAS season, with the only exception being the 6-month lead forecast, where $\bar{k}_{max}$ appears in the JFM season. For predictions of the El Niño decaying period (Fig. S7b), $\bar{k}$ values are mostly negative, and $\bar{k}_{max}$ primarily occurs in the AMJ season, except for the 1-month lead forecast, where $\bar{k}_{max}$ is observed in the OND season. For the CNN model, in 1-month lead forecasts of El Niño, $\bar{k}_{max}$ for both the growing and decaying periods falls in the OND season (Figs. S8a and S8b), indicating an absence of the SPB phenomenon; however, SPB is evident at other lead times. In the ConvTransformer model's predictions of the La Niña growing period (Fig. S7c), $\bar{k}_{max}$ occurs in the JFM season for 1- and 6-month lead times, while for other lead times, it appears in the AMJ season. For predictions of the La Niña decaying period (Fig. S7d), $\bar{k}_{max}$ is observed in the JFM and OND seasons for 6-, 9-, and 12-month lead times, and in the AMJ and JAS seasons for other lead times. For the CNN model's predictions of the La Niña growing period (Fig. S8c), $\bar{k}_{max}$ occurs in the JAS season for the 1-month lead time and in the AMJ season for the 18-month lead time, with other lead times showing $\bar{k}_{max}$ in the JFM and OND seasons. In predictions of the La Niña decaying period (Fig. S8d), $\bar{k}_{max}$ falls in the OND season for 12- and 15-month lead times, while other lead times exhibit $\bar{k}_{max}$ in the AMJ and JAS seasons. These results indicate that for La Niña forecasts, $\bar{k}_{max}$ does not consistently occur in spring, resulting in a less pronounced SPB phenomenon, whereas forecast errors for El Niño events exhibit distinct seasonal variations, confirming the presence of SPB.

In summary, the deep learning models in this study, by leveraging sea subsurface temperature data, can effectively mitigate the traditional SPB in ENSO prediction: they reduce the degradation of predictive skill caused by SPB during spring for short-to-medium lead times (1-12 months) and effectively suppress SPB signal interference in La Niña prediction. This technical advancement provides a highly promising approach for improving springtime ENSO predictive skill.

2.3. Interpreting deep learning results

In addition to effective forecasting skill, the physical mechanism underlying the ENSO prediction based on the theory of sea subsurface temperature propagation can also be unraveled. To further explore the sea subsurface temperature anomaly regions identified by deep learning model responsible for skillful prediction of ENSO, we employed Integrated Gradients (IG)⁶⁷. The Integrated Gradients (IG) method enables systematic comparison of feature attribution patterns between El Niño years and neutral years, providing a physical interpretation for the superior performance of the ConvTransformer model in El Niño event prediction. Given the distinct climatic backgrounds during the period 1871-2024, a composite analysis of attribution maps across the entire timeframe would lead to ambiguous attributions. Therefore, we divided this period into three climatic periods based on background conditions: 1871-1939 (the residual effects of the Little Ice Age and transition to modern warming)⁶⁸, 1940-1979 (mid-20th century warming slowdown/climatic regime shift period)⁶⁹, and 1980-2024 (the recent rapid warming period)⁶⁸. Composite attribution analyses were then conducted separately for El Niño events under each of these distinct climatic backgrounds. For the period 1871-1949, El Niño years (1877, 1878, 1880, 1881, 1884, 1885, 1888, 1889, 1891, 1896, 1897, 1899, 1900, 1902, 1903, 1905, 1906, 1911, 1912, 1914, 1915, 1918, 1919, 1925, 1926, 1930, 1931, 1939, 1940, 1941, 1942, 1946, 1947, 1949) were selected based on the MEI.ext (Extended Multivariate ENSO Index: NOAA Physical Sciences Laboratory) and historically reconstructed "consensus events". For 1950–2024, El Niño events (1951, 1952, 1953, 1954, 1957, 1958, 1959, 1963, 1964, 1965, 1966, 1968, 1969, 1970, 1972, 1973, 1976, 1977, 1978, 1979, 1982, 1987, 1991, 1994, 1997, 2002, 2004, 2006, 2009, 2015, 2018, 2023) were identified according to the official Oceanic Niño Index (ONI) classification from NOAA/CPC (origin.cpc.ncep.noaa.gov). Focusing on these El Niño events, this study derived composite attribution maps for the 1-month lead-time

forecast of El Niño's peak stage across the three climatic backgrounds, covering multiple depth layers (surface, 5 m, 35 m, 46 m, 110 m, 149 m, 201 m, 288 m). Regions with statistical significance ($p < 0.05$, Student's *t*-test) are highlighted in these maps (Figs. S9, S11, 3). Additionally, Figs. S10, S12, and 4 present composite ocean temperature analysis maps corresponding to the composite attribution maps across the same depth layers. These ocean temperature maps were used to verify whether the key regions identified by attribution analysis coincide with regions of ocean temperature anomalies. Examination of the composite attribution maps and ocean temperature composite maps under different climatic backgrounds reveals that the spatial distribution of ocean temperature anomalies varies across these backgrounds. However, the key regions captured by the ConvTransformer model are generally consistent with the regions of significant ocean temperature anomalies—this consistency provides a physical basis for the model's excellent performance in El Niño prediction. Given that the quality of data from 1980 onwards is higher than that of reconstructed data from 1871-1979 (e.g., ocean temperature reconstructions based on ship observations)⁷⁰, and that the climatic background since 1980 is most similar to the present day, the attribution patterns derived for this period are more reliable. Therefore, this study focuses on the analysis of attribution composite maps and ocean temperature composite maps for the 1980-2024 period. Attribution analysis results in Fig. 3 show that, when the ConvTransformer model forecasts El Niño's Phase 3, the captured key regions exhibit a distinct depth-dependent structure, which is generally consistent with the anomaly distribution of ocean temperature across different layers in Fig. 4. Specifically, at the sea surface, key regions are concentrated near 20°N in the eastern Pacific and the Niño regions (see Fig. 3a). At 5 m below the sea surface (Fig. 3b), the range of regions with high attribution values expands, focusing on the equatorial zone (125°E-175°E, 160°W-110°W)—this zone corresponds to a region of significant ocean temperature anomalies in the ocean temperature composite analysis map (Fig. 4b). From 35 m to 201 m below the sea surface (Figs. 3c-g), clear regions with high attribution values remain visible in the equatorial zone (125°E-175°E, 160°W-110°W); however, as depth increases, the range of these high-attribution regions shrinks, and they still align with the ocean temperature composite analysis maps (Figs. 4c-g). At 288 m below the sea surface, ocean temperature anomalies are weak, and the model can only capture scattered signals. These results demonstrate the ConvTransformer

model's ability to capture distinct and coherent signals during the forecast of El Niño's peak stage, validating its utility in El Niño event prediction.

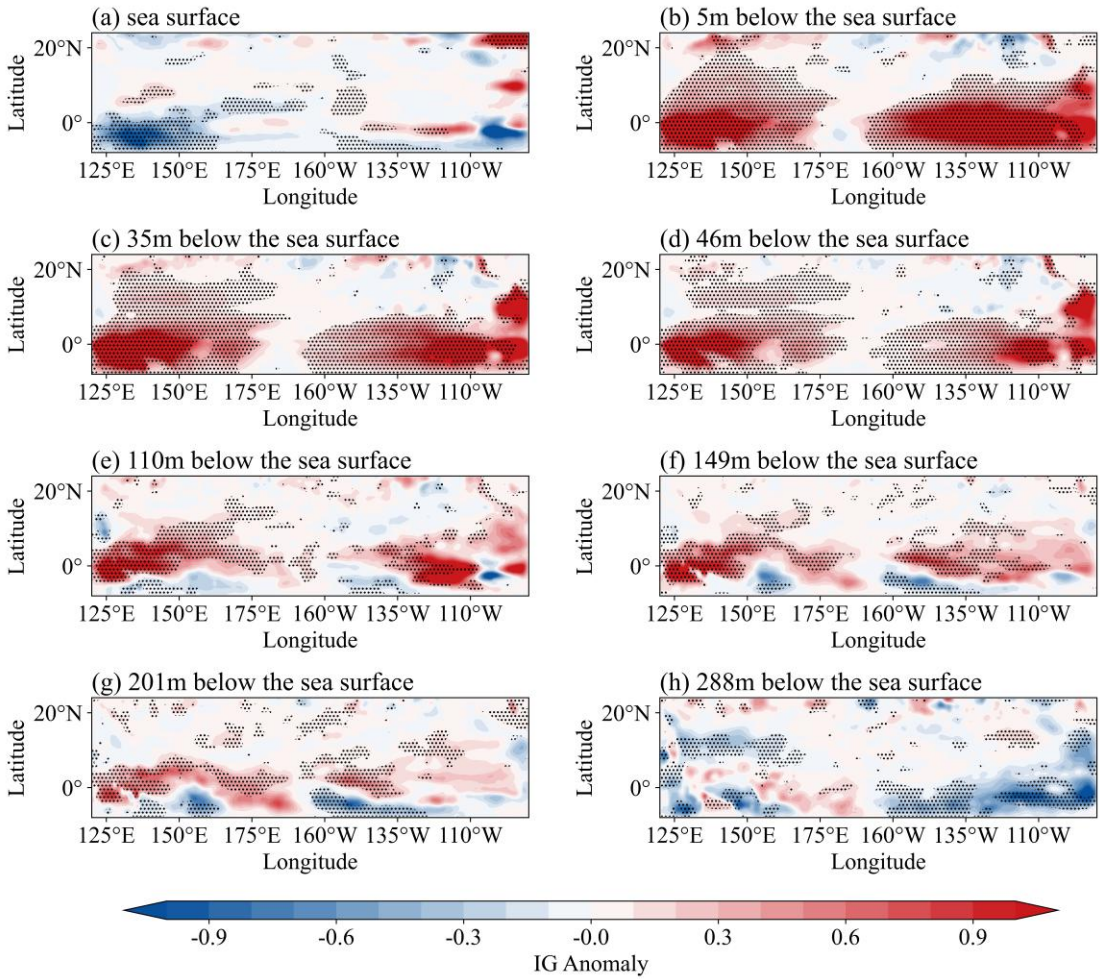


Fig. 3 Composite integrated gradient (IG) maps for El Niño events (1980-2024), illustrating key characteristics of sea surface and multi-layer subsurface fields (at 5, 35, 46, 110, 149, 201, and 288 meters) during the peak stage of El Niño events that occurred between 1980 and 2024, as forecasted by ConvTransformer model at 1-month lead time. The maps were smoothed using a Gaussian filter ($\sigma=1.0$) to reduce outlier effects. Shaded regions indicate areas where attribution values exceed the 95% confidence level based on Student's t-test, highlighting statistically significant precursors for El Niño development.

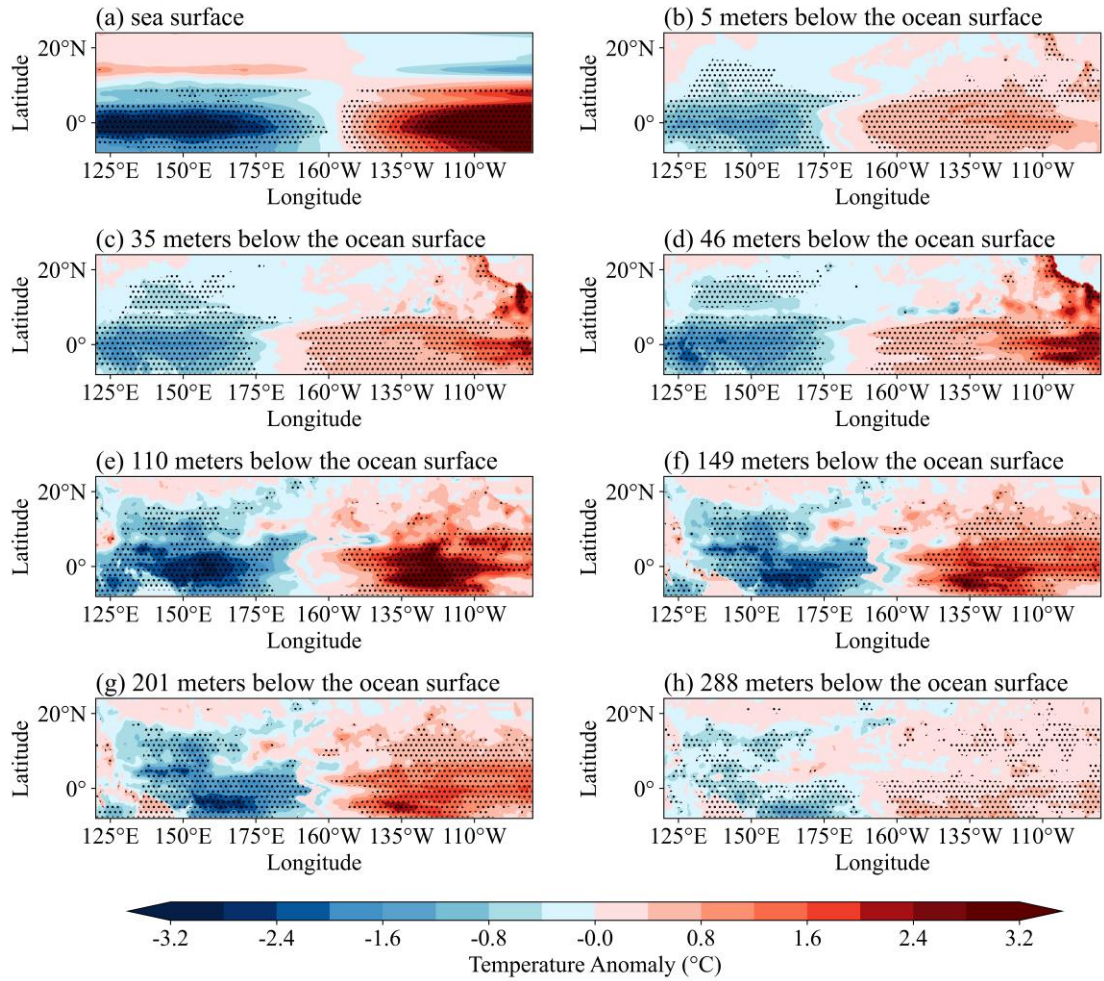


Fig. 4 Composite analysis maps of sea surface and multi-layer subsurface fields (at 5, 35, 46, 110, 149, 201, and 288 meters) during the peak stage of El Niño events (1980-2024), with the shaded areas representing regions where the values exceed the 95% confidence level.

To investigate whether the key regions captured by the ConvTransformer model differ when forecasting different stages of El Niño, Fig. S13 presents composite analysis maps of attribution values captured by the model for the 1-month lead-time forecast of El Niño's pre-peak stage, with Fig. S14 showing the corresponding composite ocean temperature analysis maps. Fig. S13 reveals that, compared to the forecast of El Niño's peak stage, the key signals captured by the model during the forecast of El Niño's pre-peak stage are weaker. Additionally, attribution analysis for La Niña events (not shown) exhibits even weaker subsurface signals, with the model focusing more on regions of significant sea surface anomalies—this pattern corresponds to the overall lower predictability of La Niña events.

Collectively, these lines of evidence demonstrate that the ConvTransformer model effectively

captures the physical precursor signals during the development of El Niño. In broad terms, the development of strong El Niño events in 1982, 1997, and 2015 is linked to ocean dynamics and the slow discharge of equatorial heat to extra-equatorial regions. The diversity of different El Niño events arises from the varying roles of noise forcing and positive/negative coupled atmosphere-ocean feedback processes⁷¹. While different predictability sources influence ENSO evolution³⁸, our study shows that model can effectively forecast El Niño events by successfully capturing the propagation of sea subsurface temperature. Likewise, Peng et al.⁷² analyzed the drivers of the 2023-2024 strong El Niño event and found, through mixed-layer heat budget analysis, that vertical advection—accounting for 62% of the warming—was the dominant contributor to the eastern Pacific warming, with thermocline feedback playing a leading role. The 2023-2024 event demonstrates that oceanic internal dynamics can independently trigger strong El Niño events without atmospheric feedback. This provides theoretical support for the model's effectiveness in forecasting strong El Niño events by capturing key oceanic dynamical processes.

2.4. Linkage between ENSO stages and subsurface patterns

The propagation of sea subsurface temperature anomalies in the tropical Pacific exhibits a well-defined loop structure, with eastward signal propagation along the equatorial waveguide (8°S-8°N) and westward propagation in the off-equatorial region (8°N-24°N), and the complete circuit period of these thermal anomalies aligns with the characteristic 2 to 7 year timescale of ENSO recurrence⁵⁰, demonstrating a robust phase-locking between subsurface dynamics and ENSO cyclicity.

Consistent with this dynamical framework, the Integrated Gradients (IG) results reveal that the attribution intensity and spatial coherence differ between the pre-peak and peak El Niño stages, implying that distinct physical processes govern different phases of event evolution. These findings motivated a stage-based partitioning approach, enabling a more detailed investigation into how subsurface temperature anomalies and coupled feedback mechanisms evolve through successive ENSO stages, and how the ConvTransformer model effectively captures these stage-dependent predictive signals.

This selected twelve El Niño years (1982, 1987, 1991, 1994, 1997, 2002, 2004, 2006, 2009, 2015, 2018, 2023) and twelve La Niña years (1983, 1984, 1988 1995, 1998, 2000, 2007, 2010,

2011, 2017, 2020, 2022) within the 1980-2024 period. The identification of El Niño and La Niña events is conducted in accordance with the official criteria established by the National Oceanic and Atmospheric Administration (NOAA) for El Niño and La Niña ([El Niño and La Niña Alert System | NOAA Climate.gov](#)). Each ENSO event was categorized into five characteristic stages (Fig. 5). The definitions of Stages 1-5 for ENSO events are as follows: Stage1 (Onset Initial Stage) specifically refers to the month when the 3-month running mean of the Niño3.4 index first exceeds 0°C (for an impending El Niño event) or falls below 0°C (for an impending La Niña event). This stage marks the initial onset of oceanic anomalies prior to the formal occurrence of the ENSO event. El Niño (or La Niña) Stage2 (Pre-peak Stage) commences when the SST continues to rise (or fall) but has not yet reached the peak of its anomalous amplitude, and the 3-month running mean of the Niño3.4 index crosses 0.5°C (or -0.5°C) for the first time. For Stage3 (Peak Stage), due to the inherent variability in the peak magnitude of the Niño3.4 index across different ENSO events (i.e., no universal quantitative threshold applies to all events), it is qualitatively defined as the month when the 3-month running mean of the Niño3.4 index reaches its maximum value during a specific El Niño or La Niña event. During Stage4 (Decay Stage), as the intensity of El Niño (or La Niña) weakens, the 3-month running mean of the Niño3.4 index drops to 0.5°C (or -0.5°C). Stage5 (Termination Stage) is the final phase of the event, during which the 3-month running mean of the Niño3.4 index gradually approaches 0°C, and the impacts of the ENSO event dissipate accordingly. This definition framework ensures a structured temporal division of the ENSO cycle across its onset, peak, and decay stages.

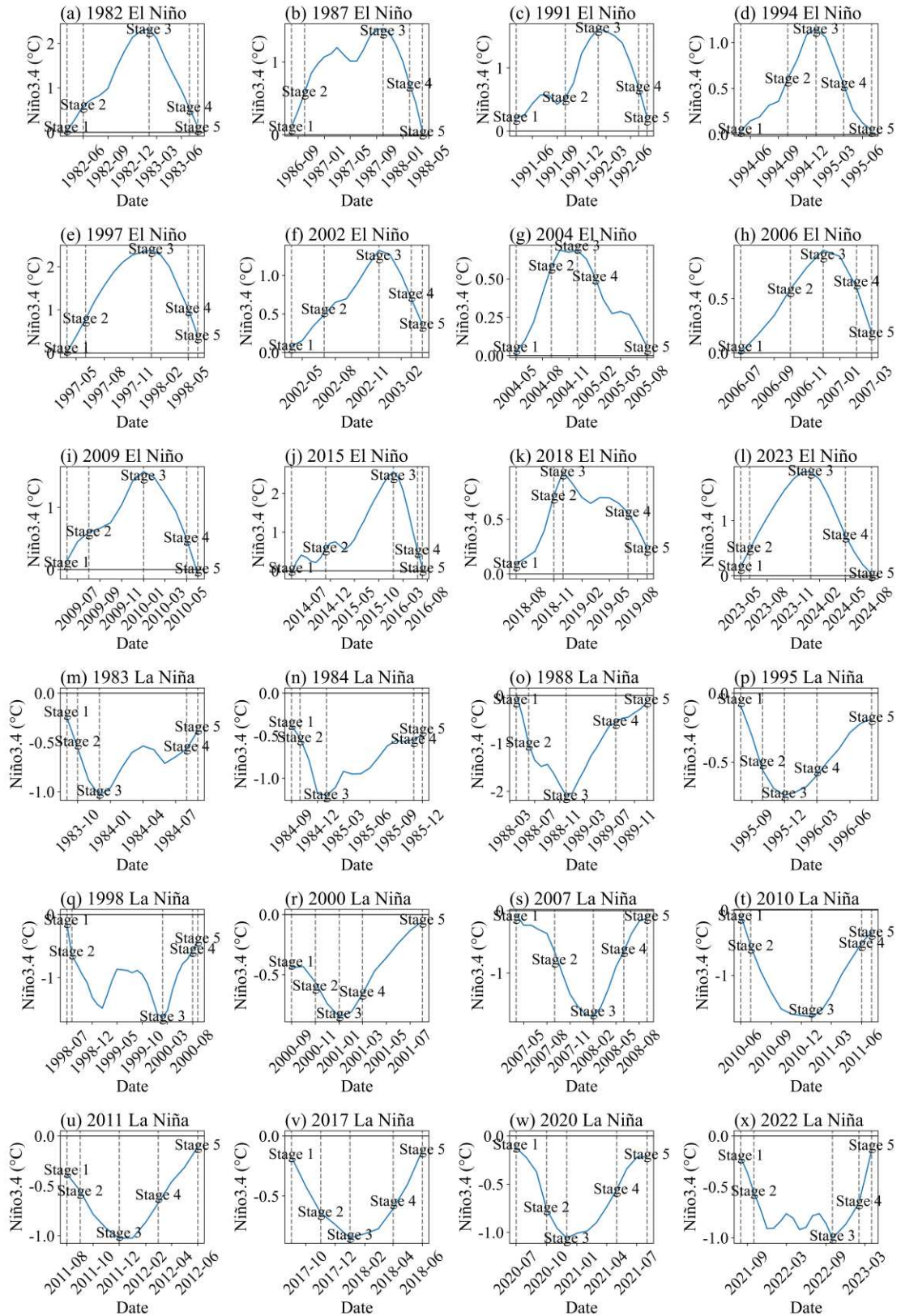


Fig. 5 El Niño years and La Niña years stage diagram of Niño3.4 index.

The distinct loop propagation of sea subsurface temperature anomalies-characterized by

eastward movement within the 8°S-8°N equatorial waveguide and westward movement in the 8°N-24°N off-equatorial band-necessitates separate analysis of these latitudinal zones. Composite analysis with statistical significance evaluated via Student's t-test ($p < 0.05$) was employed to identify characteristic propagation patterns during ENSO events. Analyzing the characteristics of sea subsurface temperature anomalies within 8°S-8°N during El Niño and La Niña events (Fig. 6) reveals that positive subsurface anomalies first emerge in the western-central Pacific (180°-140°W) during the pre-El Niño period, i.e., Stage1, gradually propagating eastward and intensifying to dominate the east-central Pacific by the peak stage, i.e., Stage3. As the event decays, these warm anomalies weaken while negative anomalies originating in the western Pacific propagate eastward, ultimately leading to the event's termination. The La Niña evolution exhibits an opposite but symmetric pattern, with cold anomalies propagating eastward during the development stage and warm anomalies emerging during decay. This latitude-dependent propagation behavior is clearly evident in the phase reversal between warm and cold events (Fig. 6).

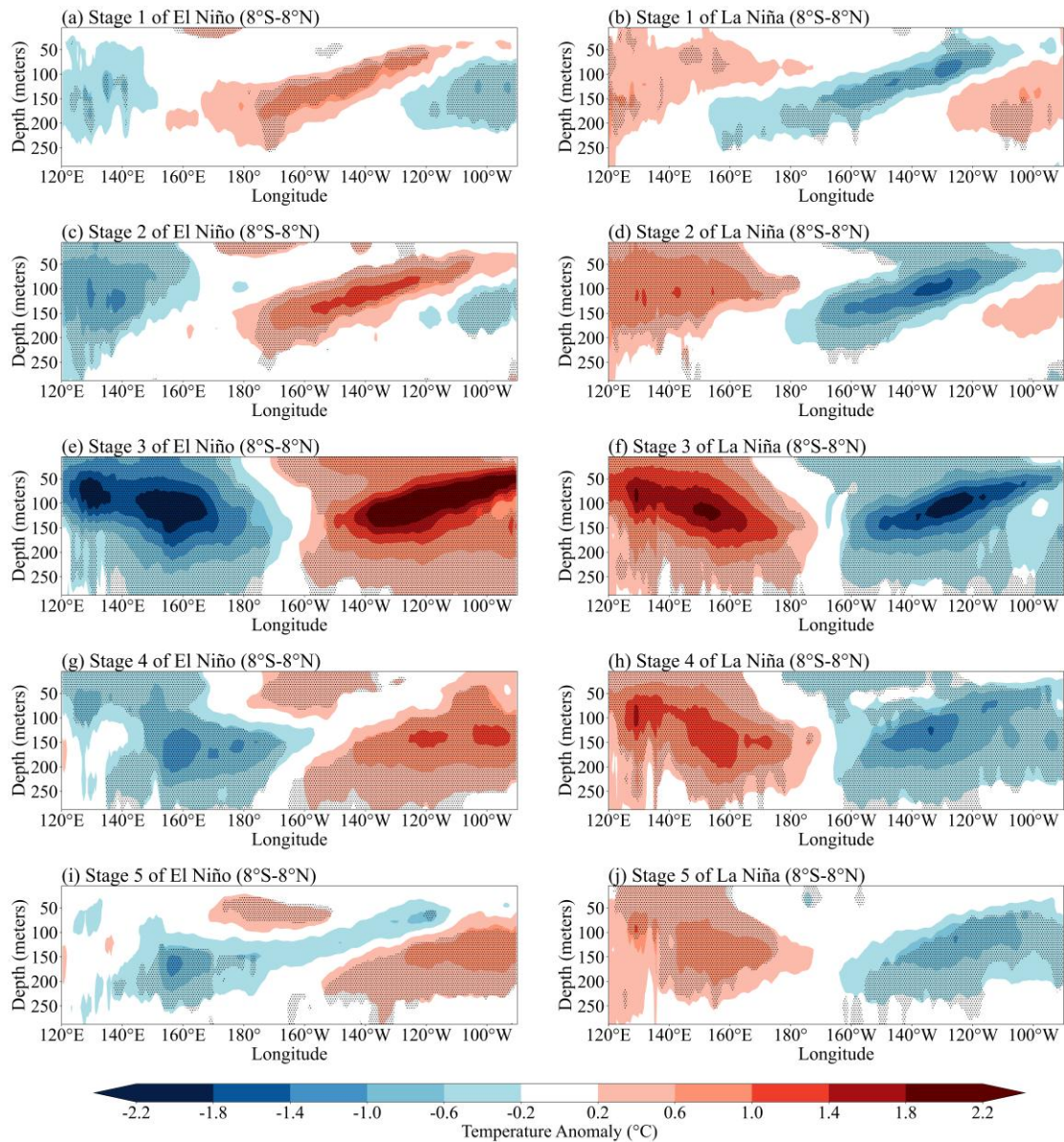


Fig. 6 Composite analysis of sea subsurface temperature anomalies for the El Niño events and the La Niña events at 8°S-8°N, with the shaded area exceeding the 95% confidence level.

The analysis of sea subsurface temperature anomalies in the off-equatorial North Pacific (8°-24°N) reveals distinct propagation characteristics during ENSO events. During El Niño development, the eastern Pacific initially exhibits negative anomalies in Stage1. As the event progresses to Stage2, the influence of positive anomalies from the equatorial region (8°S-8°N) becomes apparent, triggering a phase reversal in the 8°N-24°N band. This reversal manifests as a westward propagation of positive anomalies that intensify through Stages3 and 4, before weakening in Stage5. The La Niña composite shows an inverse propagation pattern, though

with notable asymmetry in anomaly strength. A clear shift of positive and negative phases can be seen in Fig. 7, and the positive anomalies are stronger than the negative anomalies.

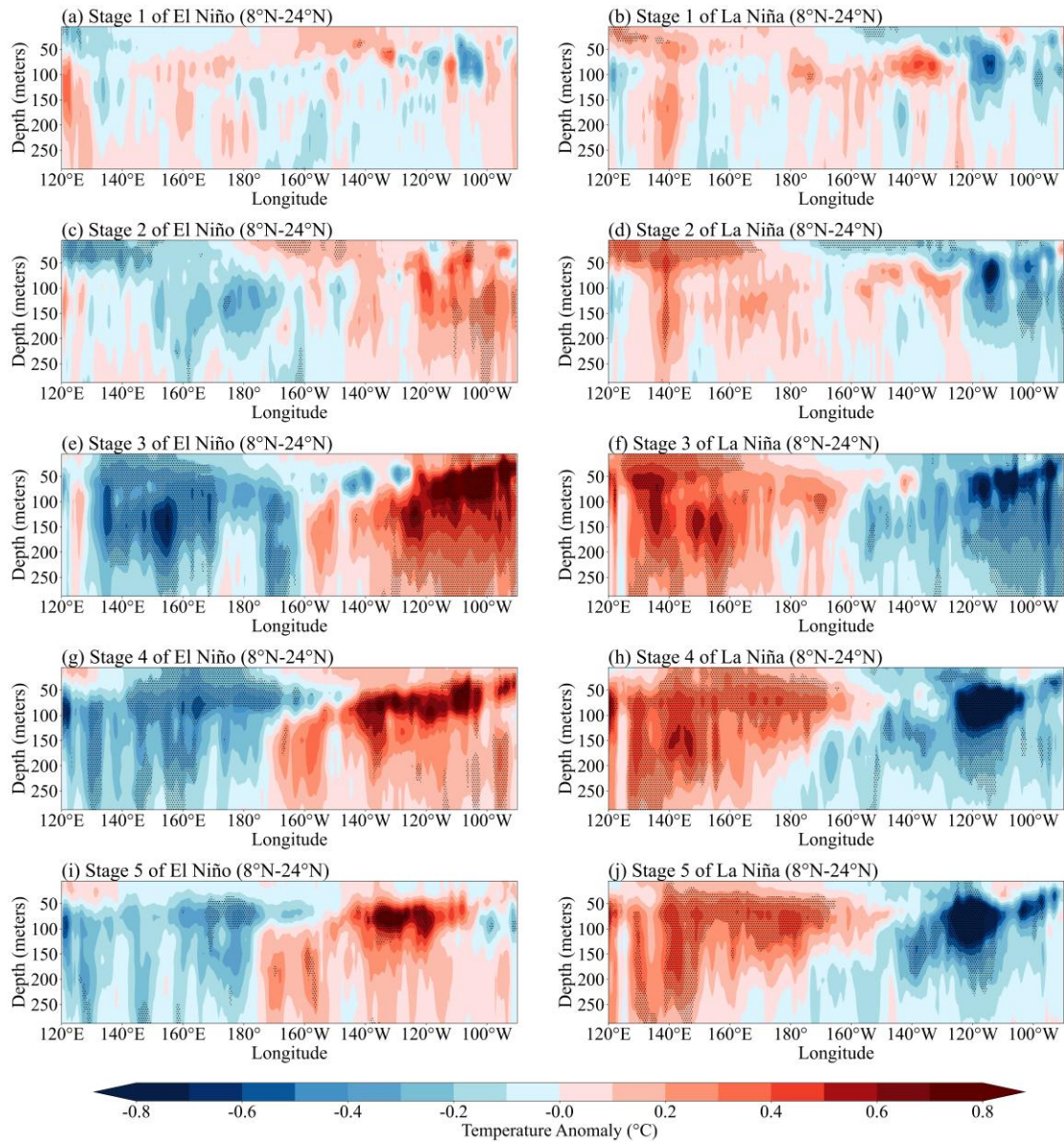


Fig. 7 Composite analysis of sea subsurface temperature anomalies for the El Niño events and the La Niña events at 8°N-24°N, with the shaded area exceeding the 95% confidence level.

Overall, subsurface oceanic signals typically exhibit stronger and more coherent precursor patterns during El Niño events compared to La Niña events, consistent with the findings of Hayashi and Jin⁶¹. This observed propagation pattern establishes subsurface thermal anomalies across the 8°S-24°N band as a physically meaningful precursor for ENSO development, with the loop mechanism serving as both a climate memory element and a potential predictor for

ENSO phase transitions. The coherent coupling between equatorial and off-equatorial subsurface processes through this propagation loop provides critical insights into ENSO's inherent predictability.

3. Discussion

Recent advances in deep learning have significantly enhanced climate forecasting capabilities, yet concerns persist regarding the physical fidelity of these data-driven approaches. To address this critical gap, our study establishes a physically-constrained framework for ENSO prediction by (i) systematically examining subsurface temperature propagation mechanisms during ENSO events, and (ii) developing two novel deep learning architectures—a CNN and ConvTransformer model—that explicitly incorporate both sea surface and subsurface temperature anomalies.

Our results demonstrate that these hybrid models successfully capture essential subsurface dynamics, enabling skillful prediction of both El Niño and La Niña events at 18-month lead times. Moreover, the ConvTransformer model identifies distinct key regions corresponding to different stages of El Niño evolution, demonstrating its ability to learn physically meaningful ENSO precursor signals while maintaining state-of-the-art predictive accuracy. This study establishes that physically-constrained deep learning models can achieve skillful ENSO predictions at 18-month lead times by leveraging sea surface temperature and sea subsurface temperature information. At the longest lead times, the streamlined CNN marginally outperforms the ConvTransformer, which we attributed to data-limited conditions. The CNN's stronger local and translation-equivariant inductive biases confer greater data efficiency when the effective sample size is constrained. By contrast, Transformers are well suited to capturing long-range dependencies but, owing to weaker priors, typically require substantially larger pretraining datasets to avoid restricted generalization or overfitting⁷³; accordingly, at shorter leads the ConvTransformer can more fully exploit nonlocal dependencies and is therefore more advantageous. Furthermore, while a simple CNN architecture already demonstrates the capability to effectively predict El Niño events at this 18-month lead time, the incorporation of a Transformer Encoder enables the model to yield more accurate forecasts of El Niño event intensity—thereby further enhancing the overall prediction skill for El Niño events. While this study's input domain (8°S-24°N, 120°E-90°W) effectively isolates equatorial-extra-equatorial

Pacific subsurface temperature interactions, it inherently limits performance by excluding extra-Pacific precursor signals. The model cannot capture out-of-domain inter-basin teleconnections, weakening its skill for ENSO events driven by extra-Pacific forcing—such as those triggered by North Tropical Atlantic warming^{74,75}, modulated by the Indian Ocean Dipole (IOD)⁷⁶, or influenced by the South Atlantic Subtropical Dipole (SASD)⁷⁷—and restricting generalizability across ENSO forcing regimes. To address these limitations, future research will expand the input domain to incorporate additional key regions. This expanded domain will enable three critical objectives: first, to evaluate whether the inclusion of extra-Pacific signals enhances overall ENSO forecasting skill, particularly for events with strong extra-Pacific forcing; second, to assess whether such an expansion can mitigate the SPB by leveraging persistent inter-basin teleconnection signals that often sustain predictability when Pacific-internal signals weaken; and third, to deepen the understanding of how multi-basin interactions collectively regulate ENSO predictability, thereby providing a more comprehensive view of the global coupled processes governing ENSO evolution. In addition, developing hybrid approaches that integrate our deep learning framework with dynamical models is to be followed up with in-depth studies. Finally, although this study has shown that the propagation mechanism of sea subsurface temperature has a significant impact on ENSO prediction, it has not yet explored how the warm and cold anomalies in the subsurface layer are generated. Investigating the generation mechanisms of subsurface thermal anomalies to complete our physical understanding of ENSO precursors is necessary.

Building on this study's findings and recent advances in ENSO prediction, future research will focus on two core priorities to advance forecasting accuracy, robustness, and interdisciplinary integration: enhancing short-term ENSO prediction and fusing climate models with DL models. First, to improve short-term ENSO prediction (lead times: 1-3 months), three key directions are proposed. Leveraging high-resolution observational/reanalysis datasets to optimize ocean-atmosphere coupled model initialization—prioritizing subsurface ocean thermal anomalies and atmospheric boundary layer processes—will reduce initial condition uncertainties, a major source of short-term forecast error. Additionally, transfer learning can enhance performance by adapting pre-trained models to ENSO-specific datasets, while refined multi-model ensembling will harness diverse model strengths to mitigate individual biases.

Second, integrating climate and DL models addresses single-method limitations. Three complementary frameworks are suggested: (i) DL models correct biases in key variables (e.g., subsurface temperature, surface wind stress) from reanalysis data to optimize climate model initialization, reducing error propagation; (ii) climate models provide physical constraints (e.g., ocean heat conservation) to DL models training/inference, resolving the "black box" issue and improving generalization for rare ENSO phases; (iii) a hybrid ensemble combines climate and DL models predictions, with dynamic weighting based on model performance in specific ENSO processes and seasonal skill variations.

Collectively, these directions will enhance short-term ENSO prediction precision and build a more physically grounded, robust forecasting system, advancing ENSO predictability across lead times and seasons.

4. Methods

4.1. Data

This study utilizes monthly sea surface temperature and sea subsurface temperature (5-300m) data from the equatorial and extra-equatorial Pacific as predictors, with the spatial domain constrained to 8°S-24°N, 120°E-90°W. This domain selection is grounded in robust physical evidence from prior literature: Qian and Hu⁵⁰ demonstrated that the propagation loop of subsurface ocean temperature anomalies in the Pacific between 8°S and 24°N is dynamically linked to ENSO variability, while Timmermann et al.⁷¹ further emphasized that oceanic heat content within this continuous equatorial–extra-equatorial Pacific region serves as the most direct and robust predictor for ENSO, providing a solid physical basis for our focus on this specific domain. The SST data are sourced from the Hadley Centre, specifically the HadISST1 (Hadley Centre Sea Ice and Sea Surface Temperature dataset version 1) product, which features an original spatial resolution of 1°×1°. Prior to model input, these SST data are spatially interpolated to a finer grid with a resolution of 0.5°×0.5°. Sea subsurface temperature data utilized in this study are derived from two versions of the Simple Ocean Data Assimilation (SODA) reanalysis, specifically SODA 2.2.4 and SODA 3.15.2, both of which feature a horizontal spatial resolution of 0.5°×0.5°. Consistent with the native variable format of SODA reanalysis datasets, we directly employed the potential temperature fields provided in these two versions. To address the difference in vertical level configurations between the two SODA

versions and ensure consistent depth matching for subsequent model training and data analysis, we performed vertical interpolation on the subsurface temperature data of SODA 2.2.4 using a cubic spline interpolation method. Following interpolation, we further standardized the interpolated SODA 2.2.4 data by calibrating its mean and variance to match those of SODA 3.15.2 at each corresponding depth level.

The dataset underwent comprehensive preprocessing to enhance signal detection. First, we removed the climatological annual cycle by calculating temporal anomalies for all variables. The time period used for trend calculation is 1981-2010. During the detrending process, the monthly ocean temperature data for each year were subtracted by the corresponding monthly mean (derived from the 1981-2010 period). Second, to better capture the characteristic propagation patterns of thermal perturbations across different vertical layers, we subsequently computed zonal anomalies from the detrended temperature fields. This two-step processing isolates the meridional perturbation component of sea surface and subsurface temperature anomalies, which exhibit significant relevance to ENSO dynamics.

The predictor variables consist of consecutive 8-month temperature fields, used to forecast the subsequent 18-month evolution of the Niño3.4 index—a core metric for ENSO quantification. Forecast lead time is defined as the temporal interval between the target prediction month and the final month of the input sequence. For the training process of the neural network, this study partitions the dataset into training, validation, and testing sets by continuous time periods. In this study, Hadley Centre observations and SODA reanalysis from 1871-1996 are used to train the neural network model. In addition, for the cross-validation analysis, this study uses Hadley Centre observations and SODA reanalysis from 1997-2010 as the validation set to evaluate the prediction skill. This temporal partitioning strategy—specifically, 1871-1996 years for training, 1997-2009 years for validation, and 2010-2024 years for testing (detailed in Table S3)—adopts non-overlapping time ranges across all subsets to eliminate data leakage. This design ensures rigorous evaluation of model generalizability while preserving temporal coherence between the training and validation processes. The SODA reanalysis provides physically consistent oceanic fields that are particularly valuable for capturing subsurface processes critical to ENSO evolution.

To evaluate the relative significance of surface and subsurface signals, we conducted

comparative forecasting experiments with three configurations: one using SST anomalies alone, one using SUBT anomalies alone, and one integrating both SST anomalies and SUBT anomalies. All three configurations utilized identical dataset partitioning (Table S3) and deep learning architectures to predict the Niño3.4 index at 1-18 forecast lead months, facilitating a direct comparison of their predictive skill and the incremental value of subsurface information. To benchmark our deep learning models, we employed retrospective forecasts (hindcasts) from Phases 1 and 2 of the North American Multi-Model Ensemble (NMME). The NMME models (detailed in Table S4) provide monthly mean SST forecasts initialized monthly from January 1980 through December 2010, with a 12-month forecast lead time. All model outputs have been processed to calculate anomalies relative to the 1981-2010 climatological mean, and standardized to a common temporal framework spanning January 1980 to December 2024, enabling direct comparison of forecast skill across initialization months.

4.2. Architecture of deep learning models

The ConvTransformer model architecture developed in this study (Fig. 8) comprises seven core components: an input layer, ConvNetwork module, Convolutional Block Attention Module (CBAM)⁷⁸, Flatten layer, Linear Projection, Transformer encoder⁷⁹, and output layer. The input layer ingests raw data, while the ConvNetwork module—composed of three convolutional layers and corresponding pooling layers—extracts spatial features via sliding kernels and compresses feature maps to mitigate overfitting and enhance translation invariance. To better capture sea subsurface temperature variation characteristics, this study incorporates a Convolutional Block Attention Module into the model. CBAM adaptively learns the channel and spatial attention weights to capture cross-dimensional feature correlations, thereby enhancing the feature representation capability of the convolutional neural network, Woo et al.⁷⁸ demonstrated improved model performance when channel dimension integration precedes spatial dimension integration. The Flatten layer converts multi-dimensional features to a 1D vector, which is dimensionally adjusted by Linear Projection before entering the Transformer encoder. The Transformer encoder consists of two stacked Transformer layers, each composed of Multi-Head Attention, Residual Connection, Layer Normalization, and Feed-Forward Network. The Multi-Head Attention, as the core component of the Transformer layer, enables parallel extraction of multi-scale features by partitioning input sequences into multiple attention

heads, thereby enhancing the model's representational capacity. In the Transformer architecture, Residual Connection and Layer Normalization are typically employed in tandem: Residual Connection preserves information through multi-layer transformations and maintain gradient flow, while Layer Normalization stabilizes activation value distributions after each sub-layer (e.g., Multi-Head Attention, Feed-Forward Network) to accelerate training. The Feed-Forward Network, typically structured as two fully connected layers with a nonlinear activation function in between, further processes the Multi-Head Attention outputs to capture complex feature interactions. In the ConvTransformer model, the Transformer encoder processes feature vectors extracted by the ConvNetwork module and projected to a compatible dimension, leveraging stacked self-attention mechanisms to model long-range dependencies across sequential features. These processed features are then fed into a fully connected layer to generate Niño3.4 index predictions.

In contrast, the companion CNN model adheres to a conventional architecture optimized for spatial feature extraction. Its six components include: input layer, convolutional layer, pooling layer, Convolutional Block Attention Module (CBAM), fully connected layer and output layer. This architecture balances feature extraction efficiency (three convolutional layers) with attention-guided refinement (CBAM), enabling the model to discern subtle precursors of ENSO events from complex oceanic data.

Both models utilize ocean temperature anomalies as input data, with the target output being the 3-month running mean of the Niño3.4 index. Both models are configured with 136 input channels to ensure sufficient capacity for extracting complex spatiotemporal features of ocean thermal variations. To mitigate overfitting during model training, a Dropout regularization technique is employed with a dropout rate of 0.5. The total number of trainable parameters for the ConvTransformer model and the CNN model is 34,255,368 and 10,644,148, respectively.

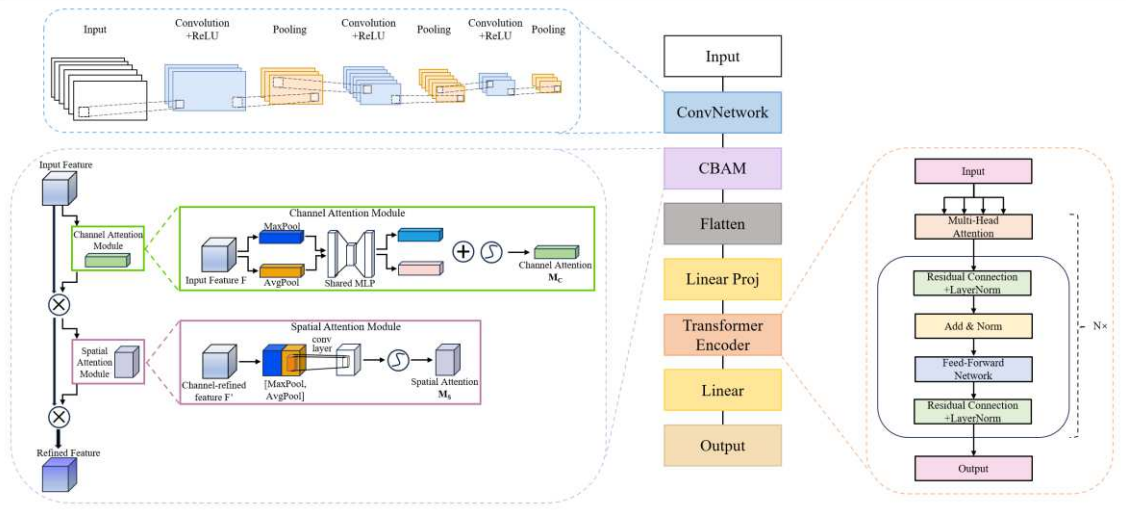


Fig. 8 Overall model architecture of ConvTransformer.

4.3. Model training strategies

The model was trained using the AdamW optimizer⁸⁰. Quantile Loss (QL)⁸¹ was adopted as the loss function, and its formulation is provided in Equation (1):

$$QuantileLoss = \frac{1}{N} \sum_{i=1}^N \Pi_{\hat{y}_i \geq y_i} (1 - \gamma) |y_i - \hat{y}_i| + \Pi_{\hat{y}_i < y_i} \gamma |y_i - \hat{y}_i| \quad (1)$$

where y_i is the actual target value, $\hat{y}_i$ is the predicted value of the model, and γ denotes the quantile level, which typically ranges between 0 and 1. In this study, the quantile loss is adapted: the quantile level is set to 0.8 for positive true values and 0.3 for negative true values. Furthermore, the batch size is configured as 256, representing the number of samples per training iteration. The learning rate, a hyperparameter controlling the speed at which the model adjusts weights based on the loss function's gradient during training, was set to 0.001. To mitigate the risk of overfitting, we adopted a model selection strategy that prioritizes the checkpoint corresponding to the lowest validation loss during the entire training process.

In this study, both the proposed ConvTransformer model and CNN model were trained as single unified frameworks rather than separate models for individual lead times or initialization months. Specifically, for each model, we used a consistent training framework where the input consists of consecutive 8-month sequences of ocean temperature anomalies, and the output directly predicts the Niño3.4 index across all target lead times (1 to 18 months). This unified design eliminates the need to train distinct models for each lead time; instead, the model learns

to capture the time-varying relationships between the 8-month input ocean temperature sequences and Niño3.4 variability at different forecast horizons through its internal hierarchical feature extraction. Regarding initialization months, the unified model is further designed to cover all calendar months as initial conditions. During the training phase, we constructed the input-output dataset by sliding the 1-month input window across every month of the historical record, ensuring that the model is exposed to initialization signals from all seasons. This approach avoids the limitation of training separate models for individual initialization months and enables the model to learn seasonally varying precursor signals that contribute to ENSO predictability across different initial conditions.

4.4. Forecast verification methods

The temporal anomaly correlation coefficient (ACC) and root-mean-square error (RMSE) are utilized to validate the performance of the model. They are computed as a function of the forecast lead month t :

$$ACC(t) = \frac{1}{12} \sum_{m=1}^{12} \frac{\sum_{y=s}^e (T_{y,m} - \overline{T_m}) (F_{y,m,t} - \overline{F_{m,t}})}{\sqrt{\sum_{y=s}^e (T_{y,m} - \overline{T_m})^2 \sum_{y=s}^e (F_{y,m,t} - \overline{F_{m,t}})^2}} \quad (2)$$

$$RMSE(t) = \frac{1}{12} \sum_{m=1}^{12} \sqrt{\frac{1}{N} \sum_{y=1}^N (F_{y,m,t} - T_{y,m})^2} \quad (3)$$

Where T and F denote the measured and forecast values, respectively; m represents target months (ranging from 1 to 12), and t is the forecast lead months; $\overline{T_m}$ and $\overline{F_{m,t}}$ denote the temporal climatology, respectively; the label y denotes the forecast target year, and N is the number of target years to predict. For different forecast lead times, the ACC and RMSE are typically calculated separately for each calendar month, and the results across all months are then averaged to derive the all-season ACC scores and RMSE scores corresponding to each lead time. This approach is designed to avoid overestimating forecast performance caused by the phase-locking behavior of the ENSO—a phenomenon that leads to significant seasonal variations in both climate variability and predictability.

When the research focus is on ENSO extreme events, the evaluation using full-sample RMSE tends to be dominated by non-extreme phases (which have a larger sample size), thereby obscuring the model improvements achieved during extreme phases. To address this issue, we

reported lead-wise verification over forecast lead months using two extremes-aligned metrics. First, a conditional RMSE (Equation 4) is computed only over extreme months for each lead to specifically evaluate the model performance during ENSO extreme events, where extremes are defined lead-wise by the fixed threshold. Second, to highlight extremes within the all-sample verification, we compute weighted RMSE (Equation 5) with larger sample weights assigned to extreme months and smaller weights to non-extreme months; to ensure comparability with the unweighted metrics, the weights at each lead time were normalized such that their sum equaled the number of valid samples for that lead.

$$\text{RMSE}^{(\text{ext})}(t) = \sqrt{\frac{1}{|\Omega_t|} \sum_{i \in \Omega_t} (F_{i,t} - T_{i,t})^2}, \Omega_t = \{i: |T_{i,t}| > \tau\} \quad (4)$$

$$\text{RMSE}^{(w)}(t) = \frac{1}{12} \sum_{m=1}^{12} \sqrt{\frac{1}{\sum_{i=1}^{N_{t,m}} \tilde{w}_{i,t,m}} \sum_{i=1}^{N_{t,m}} \tilde{w}_{i,t,m} (F_{i,t} - T_{i,t})^2} \quad (5)$$

Let $T_{i,t}$ and $F_{i,t}$ denote the observed and predicted values for sample i at forecast lead t . For the conditional RMSE, $|\Omega_t|$ denotes the number of extreme samples at lead time t . For the weighted RMSE, $N_{t,m}$ denotes the total number of samples for month m at forecast lead time t ; the weight $\tilde{w}_{i,t,m}$ was assigned as w_{ext} when $|T_{i,t}| > \tau$, and as w_{non} otherwise. When calculating the weighted RMSE across different forecast lead times, we first computed the weighted RMSE for each calendar month individually, followed by averaging these monthly values to derive the final result for each lead time. In this study, extremes were defined using a fixed threshold τ , with $\tau = 1.0^\circ\text{C}$ for the Niño3.4 index. This threshold is commonly adopted to distinguish moderate and stronger ENSO events, thereby ensuring that a broader range of extreme cases is captured. We set $w_{\text{ext}} = 3.0$ and $w_{\text{non}} = 1.0$, meaning that extreme samples were given three times the weight of non-extreme samples (while normalization of the weights ensured comparability across metrics). To assess the sensitivity of the results to the weighting design, this study further compared the weighted RMSE under three weighting schemes, with $w_{\text{ext}} = 2, 3$, and 5 (Fig. S15). Persistence and MME were also included as reference benchmarks. Here, persistence is defined as a forecast in which, for each sample, the Niño3.4 index at the initialization time is held constant for all subsequent lead times. The MME is

defined as the equally weighted mean of forecasts from the seven NMME member models. The results show that, although different weighting schemes alter the absolute values of weighted RMSE for each model, they do not change the primary relative relationships among the models. This indicates that the main conclusions regarding differences in model performance are not contingent on a single choice of weight, but are instead robust across a range of weighting settings.

The conditional RMSE isolates forecast errors during strong ENSO phases, avoiding the dilution effect from neutral phases and better reflecting skill in predicting extremes⁸², whereas the weighted RMSE aligns forecast verification with the research focus on extremes and mitigates the tendency of RMSE to be dominated by high-variance non-extreme samples⁸³. Using both jointly enables a more comprehensive evaluation of ENSO forecast skill and a clearer identification of strengths and weaknesses in model improvements.

To assess the statistical significance of forecast skill differences among different forecasting schemes, this study applied a bootstrap resampling approach to test the significance of ACC, RMSE, conditional RMSE, and weighted RMSE. The significance assessment was conducted against three types of reference benchmarks: (i) the random forecast benchmark; (ii) the persistence benchmark; and (iii) the MME benchmark. For ACC, an improvement in skill was regarded as statistically significant only when the model ACC was significantly higher than that of the corresponding benchmark. For RMSE, conditional RMSE, and weighted RMSE, model performance was considered superior only when the corresponding error metric was significantly lower than that of the reference benchmark. The random forecast benchmark was not constructed by generating arbitrary random number sequences. Instead, for each forecast lead time and initialization month, the observed Niño3.4 index series was randomly permuted, thereby preserving the marginal distribution characteristics of the samples for that specific month and lead time while disrupting the correspondence between the predictions and the observations.

To quantitatively evaluate forecasting effectiveness, three evaluation indexes were employed: (i) Hit Rate (Sensitivity or Recall, HR), defined as the ratio of hits to the sum of hits and misses (Equation 6). (ii) False Alarm Ratio (FAR), defined as the ratio of false alarms to the sum of hits plus false alarms (Equation 7). (iii) Critical Success Index (CSI), defined as the ratio of hits

to the sum of hits, misses, and false alarms⁸⁴ (Equation 8). The thresholds for identifying ENSO events in this study were set as ± 0.5 ⁸⁵.

$$HR = \frac{target_is_target}{target_is_target + target_is_nontarget} \quad (6)$$

$$FAR = \frac{nontarget_is_target}{target_is_target + nontarget_is_target} \quad (7)$$

$$CSI = \frac{target_is_target}{target_is_target + target_is_nontarget + nontarget_is_target} \quad (8)$$

4.5. Seasonal growth rate of forecast error

The Spring Predictability Barrier (SPB) phenomenon during the growing and decaying periods of forecasted El Niño and La Niña events was investigated using the seasonal growth rate (SGR) of forecast errors. For a target month t , the forecast error $E(t)$ was defined as the difference between the model's forecast value F_t and the observed value T_t , expressed as follows:

$$E(t) = \sqrt{(F_t - T_t)^2} \quad (9)$$

The error growth rate for the target month t was calculated as the difference between the forecast error in month $t + 1$ and that in month t . The calendar year was divided into four seasons: JFM (January–February–March), AMJ (April–May–June), JAS (July–August–September), and OND (October–November–December). The seasonal growth rate of forecast errors $\bar{k}$ was defined as the average of the monthly error growth rates within each season. The SPB phenomenon was considered to occur if the maximum seasonal error growth rate $\bar{k}_{max}$ occurred in either the AMJ or JAS season^{86,87}. Following the definition proposed by Duan and Wei⁸⁷, the growing period of an El Niño (La Niña) event was delineated as the period from the first month when the Niño3.4 index exceeded -0.5° (fell below 0.5°) to the month when the index reached its peak. The decaying period was defined as the period from the peak month to the first month when the Niño3.4 index dropped below -0.5° (rose above 0.5°).

4.6. Integrated Gradients for physical precursor identification

IG is an explainable AI technique that quantifies feature importance. The attribution values are calculated as follows:

$$IntegratedGrads_i(x) = (x_i - x'_i) \times \sum_{k=1}^m \frac{\partial F\left(x' + \frac{k}{m} \times (x - x')\right)}{\partial x_i} \times \frac{1}{m} \quad (10)$$

where m is the number of steps in the Riemann approximation of the integral; i is the index of the input pixels to the model; x' is the baseline chosen for comparing the results in this study,

which is set to be the same zero vector as the shape of the model inputs; $IntegratedGrads_i(x)$ is the calculated attribution value. The magnitude of attribution values can indicate which areas are sensitive and have a significant impact on the predictive ability of the Niño3.4 index.

Data availability

NMME data downloaded from <https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/>.

Sea surface temperature data was obtained from the Hadley Centre, downloaded from [Met Office Hadley Centre observations datasets](#).

Sea subsurface temperature data was used from SODA, downloaded from [SODA Reanalysis Project](#).

Acknowledgements

X.Z. and Y.W. were supported by National Natural Science Foundation of China (Grant No. 42275164).

J.W. was supported by grants from the Ministry of Education of Singapore (#Tier2 MOE-T2EP50124-0018 and #Tier3 MOET32022-0006). This work comprises EOS contribution number 760.

Author contributions

Y.W.: Writing - review & editing, Writing - original draft, Visualization, Validation, Methodology, Investigation, Formal analysis.

X.Z.: Writing - review & editing, Writing - original draft, Visualization, Validation, Supervision, Project administration, Investigation, Formal analysis, Conceptualization.

J.W.: Writing - review & editing, Writing - original draft, Visualization, Validation, Supervision, Project administration, Investigation, Formal analysis, Conceptualization.

L.Z.: Writing - review & editing, Visualization, Validation, Resources, Formal analysis.

Y.L.: Writing - review & editing, Methodology, Visualization, Validation, Data curation, Conceptualization.

Y.J.: Writing - review & editing, Methodology, Visualization, Validation, Data curation, Conceptualization.

S.Z.: Writing - review & editing, Visualization, Validation.

D.K.: Writing - review & editing, Visualization, Validation.

K.F.: Writing - review & editing, Visualization, Validation.

Y.C.: Writing - review & editing, Visualization, Validation.

L.Q.Z.: Writing - review & editing, Visualization, Validation.

Competing interests

The authors declare no competing interests.

References

- 1 Philander, S. G. El Niño, La Niña, and the southern oscillation. *International geophysics series* **46**, 9-57, doi:[https://doi.org/10.1016/S0074-6142\(08\)60172-2](https://doi.org/10.1016/S0074-6142(08)60172-2) (1990).
- 2 McCreary, J. A linear stratified ocean model of the equatorial undercurrent. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical* **298**, 603-635 (1981).
- 3 Philander, S., Hurlin, W. & Seigel, A. Simulation of the seasonal cycle of the tropical Pacific Ocean. *Journal of Physical Oceanography* **17**, 1986-2002 (1987).
- 4 Neelin, J. *et al.* Tropical air-sea interaction in general circulation models. *Climate Dynamics* **7**, 73-104, doi:<https://doi.org/10.1007/BF00209610> (1992).
- 5 Fraedrich, K. An ENSO impact on Europe? *Tellus A* **46**, 541-552, doi:<https://doi.org/10.1034/j.1600-0870.1994.00015.x> (1994).
- 6 Hu, K., Huang, G., Huang, P., Kosaka, Y. & Xie, S.-P. Intensification of El Niño-induced atmospheric anomalies under greenhouse warming. *Nature Geoscience* **14**, 377-382, doi:<https://doi.org/10.1038/s41561-021-00730-3> (2021).
- 7 Geng, X., Zhao, J. & Kug, J.-S. ENSO-driven abrupt phase shift in North Atlantic oscillation in early January. *npj Climate and Atmospheric Science* **6**, 80, doi:<https://doi.org/10.1038/s41612-023-00414-2> (2023).
- 8 Tang, S. *et al.* Recent changes in ENSO's impacts on the summertime circumglobal teleconnection and mid-latitude extremes. *Nature Communications* **16**, 646, doi:<https://doi.org/10.1038/s41467-025-55925-8> (2025).
- 9 Bayr, T. & Latif, M. ENSO atmospheric feedbacks under global warming and their relation to mean-state changes. *Climate Dynamics* **60**, 2613-2631, doi:<https://doi.org/10.1007/s00382-022-06454-3>

- (2023).
- 10 Wang, G., Cai, W., Santoso, A. & Yang, K. Southern Ocean heat buffer constrained by present-day ENSO teleconnection. *npj Climate and Atmospheric Science* **7**, 181, doi:<https://doi.org/10.1038/s41612-024-00731-0> (2024).
- 11 Chen, Z. *et al.* The impact of global warming on ENSO from the perspective of objective signals. *Atmospheric Research* **299**, 107176, doi:<https://doi.org/10.1016/j.atmosres.2023.107176> (2024).
- 12 Lu, Z. *et al.* Increased frequency of multi-year El Niño–Southern Oscillation events across the Holocene. *Nature Geoscience* **18**, 337–343, doi:<https://doi.org/10.1038/s41561-025-01670-y> (2025).
- 13 Dominiak, S. & Terray, P. Improvement of ENSO prediction using a linear regression model with a southern Indian Ocean sea surface temperature predictor. *Geophysical Research Letters* **32**, L18702, doi:<https://doi.org/10.1029/2005GL023153> (2005).
- 14 Zhang, R.-H., Zhou, L., Gao, C. & Tao, L. A transformer-based coupled ocean-atmosphere model for ENSO studies. *Science Bulletin* **69**, 2323–2327, doi:10.1016/j.scib.2024.04.048 (2024).
- 15 Eyring, V. *et al.* Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958, doi:<https://doi.org/10.5194/gmd-9-1937-2016> (2016).
- 16 Fraedrich, K. & Smith, N. R. Combining Predictive Schemes in Long-Range Forecasting. *Journal of Climate* **2**, 291–294, doi:<http://www.jstor.org/stable/26195818> (1989).
- 17 Metzger, S., Latif, M. & Fraedrich, K. Combining ENSO Forecasts: A Feasibility Study. *Monthly Weather Review* **132**, 456–472, doi:[https://doi.org/10.1175/1520-0493\(2004\)132<0456:CEFAFS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0456:CEFAFS>2.0.CO;2) (2004).
- 18 Zebiak, S. E. & Cane, M. A. A Model El Niño–Southern Oscillation. *Monthly Weather Review* **115**, 2262–2278, doi:[https://doi.org/10.1175/1520-0493\(1987\)115<2262:AMENO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<2262:AMENO>2.0.CO;2) (1987).
- 19 Xue, Y., Leetmaa, A. & Ji, M. ENSO Prediction with Markov Models: The Impact of Sea Level. *Journal of Climate* **13**, 849–871, doi:[https://doi.org/10.1175/1520-0442\(2000\)013<0849:EPWMMT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<0849:EPWMMT>2.0.CO;2) (2000).
- 20 Clarke, A. J. & Van Gorder, S. Improving El Niño prediction using a space–time integration of Indo-Pacific winds and equatorial Pacific upper ocean heat content. *Geophysical Research Letters* **30**, 1399, doi:<https://doi.org/10.1029/2002GL016673> (2003).
- 21 Ruiz, J. E., Cordery, I. & Sharma, A. Integrating ocean subsurface temperatures in statistical ENSO

- forecasts. *Journal of Climate* **18**, 3571-3586, doi:<https://doi.org/10.1175/JCLI3477.1> (2005).
- 22 Drosowsky, W. Statistical prediction of ENSO (Nino 3) using sub-surface temperature data. *Geophysical Research Letters* **33**, L03710, doi:<https://doi.org/10.1029/2005GL024866> (2006).
- 23 Luo, J.-J., Masson, S., Behera, S. K. & Yamagata, T. Extended ENSO predictions using a fully coupled ocean-atmosphere model. *Journal of Climate* **21**, 84-93, doi:<https://doi.org/10.1175/2007JCLI1412.1> (2008).
- 24 Lima, C. H. R., Lall, U., Jebara, T. & Barnston, A. G. Statistical Prediction of ENSO from Subsurface Sea Temperature Using a Nonlinear Dimensionality Reduction. *Journal of Climate* **22**, 4501-4519, doi:<https://doi.org/10.1175/2009JCLI2524.1> (2009).
- 25 Ren, H.-L. *et al.* Seasonal predictability of winter ENSO types in operational dynamical model predictions. *Climate Dynamics* **52**, 3869-3890, doi:<https://doi.org/10.1007/s00382-018-4366-1> (2019).
- 26 Wang, X., Slawinska, J. & Giannakis, D. Extended-range statistical ENSO prediction through operator-theoretic techniques for nonlinear dynamics. *Scientific Reports* **10**, 2636, doi:<https://doi.org/10.1038/s41598-020-59128-7> (2020).
- 27 Flügel, M. & Chang, P. Does the predictability of ENSO depend on the seasonal cycle? *Journal of the atmospheric sciences* **55**, 3230-3243, doi:[https://doi.org/10.1175/1520-0469\(1998\)055<3230:DTPOED>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<3230:DTPOED>2.0.CO;2) (1998).
- 28 Ham, Y.-G., Kim, J.-H. & Luo, J.-J. Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568-572, doi:<https://doi.org/10.1038/s41586-019-1559-7> (2019).
- 29 Ye, M. *et al.* Multi-Year ENSO Forecasts Using Parallel Convolutional Neural Networks With Heterogeneous Architecture. *Frontiers in Marine Science* **8**, 717184, doi:<https://doi.org/10.3389/fmars.2021.717184> (2021).
- 30 Gupta, M., Kodamana, H. & Sandeep, S. Prediction of ENSO Beyond Spring Predictability Barrier Using Deep Convolutional LSTM Networks. *Ieee Geoscience and Remote Sensing Letters* **19**, 1-5, doi:10.1109/LGRS.2020.3032353 (2022).
- 31 Wang, Y., Zhang, Y. & Wang, G.-G. Forecasting ENSO using convolutional LSTM network with improved attention mechanism and models recombined by genetic algorithm in CMIP5/6. *Information Sciences* **642**, 119106, doi:<https://doi.org/10.1016/j.ins.2023.119106> (2023).
- 32 Lyu, P. *et al.* ResoNet: Robust and Explainable ENSO Forecasts with Hybrid Convolution and

Transformer Networks. *Advances in Atmospheric Sciences* **41**, 1289-1298,
doi:<https://doi.org/10.1007/s00376-024-3316-6> (2024).

33 Chen, Y. *et al.* Combined dynamical-deep learning ENSO forecasts. *Nature Communications* **16**, 1-
8, doi:<https://doi.org/10.1038/s41467-025-59173-8> (2025).

34 Frauen, C. & Dommenges, D. Influences of the tropical Indian and Atlantic Oceans on the
predictability of ENSO. *Geophysical Research Letters* **39**, L02706,
doi:<https://doi.org/10.1029/2011GL050520> (2012).

35 Joshi, M. K., Abid, M. A. & Kucharski, F. The Role of an Indian Ocean Heating Dipole in the ENSO
Teleconnection to the North Atlantic European Region in Early Winter during the Twentieth Century
in Reanalysis and CMIP5 Simulations. *Journal of Climate* **34**, 1047-1060,
doi:<https://doi.org/10.1175/JCLI-D-20-0269.1> (2021).

36 Fan, L. & Meng, X. The Asymmetric Predictive Power of Indian Ocean Dipole for Subsequent Year's
ENSO: Role of Atlantic Ocean as an Intermediary. *Geophysical Research Letters* **50**,
e2023GL105525, doi:<https://doi.org/10.1029/2023GL105525> (2023).

37 Jin, Y. *et al.* The Indian Ocean Weakens the ENSO Spring Predictability Barrier: Role of the Indian
Ocean Basin and Dipole Modes. *Journal of Climate* **36**, 8331-8345,
doi:<https://doi.org/10.1175/JCLI-D-22-0800.1> (2023).

38 Zhao, S. *et al.* Explainable El Niño predictability from climate mode interactions. *Nature* **630**, 891–
898, doi:<https://doi.org/10.1038/s41586-024-07534-6> (2024).

39 Thompson, N. C., Greenewald, K., Lee, K. & Manso, G. F. Deep Learning's Diminishing Returns:
The Cost of Improvement is Becoming Unsustainable. *IEEE Spectrum* **58**, 50-55,
doi:10.1109/MSPEC.2021.9563954 (2021).

40 Shen, C. *et al.* Differentiable modelling to unify machine learning and physical models for
geosciences. *Nature Reviews Earth & Environment* **4**, 552-567, doi:<https://doi.org/10.1038/s43017-023-00450-9> (2023).

41 Bjerknes, J. ATMOSPHERIC TELECONNECTIONS FROM THE EQUATORIAL PACIFIC.
Monthly Weather Review **97**, 163-172, doi:[https://doi.org/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).

42 Wyrski, K. El Niño—The Dynamic Response of the Equatorial Pacific Ocean to Atmospheric Forcing.
Journal of Physical Oceanography **5**, 572-584, doi:<https://doi.org/10.1175/1520->

- 0485(1975)005<0572:ENTDRO>2.0.CO;2 (1975).
- 43 Suarez, M. J. & Schopf, P. S. A Delayed Action Oscillator for ENSO. *Journal of Atmospheric Sciences* **45**, 3283-3287, doi:[https://doi.org/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).
- 44 Jin, F.-F. An Equatorial Ocean Recharge Paradigm for ENSO. Part I: Conceptual Model. *Journal of the Atmospheric Sciences* **54**, 811-829, doi:[https://doi.org/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).
- 45 Jin, F.-F. An Equatorial Ocean Recharge Paradigm for ENSO. Part II: A Stripped-Down Coupled Model. *Journal of the Atmospheric Sciences* **54**, 830-847, doi:[https://doi.org/10.1175/1520-0469\(1997\)054<0830:AEORPF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1997)054<0830:AEORPF>2.0.CO;2) (1997).
- 46 Rasmusson, E. M. & Carpenter, T. H. Variations in Tropical Sea Surface Temperature and Surface Wind Fields Associated with the Southern Oscillation/El Niño. *Monthly Weather Review* **110**, 354-384, doi:[https://doi.org/10.1175/1520-0493\(1982\)110<0354:VITSST>2.0.CO;2](https://doi.org/10.1175/1520-0493(1982)110<0354:VITSST>2.0.CO;2) (1982).
- 47 Neelin, J. D. *et al.* ENSO theory. *Journal of Geophysical Research: Oceans* **103**, 14261-14290, doi:<https://doi.org/10.1029/97JC03424> (1998).
- 48 Gu, D. & Philander, S. G. H. Interdecadal Climate Fluctuations That Depend on Exchanges Between the Tropics and Extratropics. *Science* **275**, 805-807, doi:10.1126/science.275.5301.805 (1997).
- 49 Lysne, J., Chang, P. & Giese, B. Impact of the extratropical Pacific on equatorial variability. *Geophysical Research Letters* **24**, 2589-2592, doi:<https://doi.org/10.1029/97GL02751> (1997).
- 50 Qian, W.-H. & Hu, H.-R. Signal propagations and linkages of subsurface temperature anomalies in the tropical Pacific and Indian Ocean. *Progress in Natural Science* **15**, 804-809, doi:<https://doi.org/10.1080/10020070512331342950> (2005).
- 51 Meinen, C. S. & McPhaden, M. J. Observations of Warm Water Volume Changes in the Equatorial Pacific and Their Relationship to El Niño and La Niña. *Journal of Climate* **13**, 3551-3559, doi:[https://doi.org/10.1175/1520-0442\(2000\)013<3551:OOWWVC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<3551:OOWWVC>2.0.CO;2) (2000).
- 52 McPhaden, M. J. Tropical Pacific Ocean heat content variations and ENSO persistence barriers. *Geophysical research letters* **30**, 1480, doi:<https://doi.org/10.1029/2003GL016872> (2003).
- 53 Ballester, J., Bordoni, S., Petrova, D. & Rodo, X. On the dynamical mechanisms explaining the western Pacific subsurface temperature buildup leading to ENSO events. *Geophysical Research Letters* **42**, 2961-2967, doi:10.1002/2015GL063701 (2015).

- 54 Yuan, X., Jin, F.-F. & Zhang, W. A Concise and Effective Expression Relating Subsurface Temperature to the Thermocline in the Equatorial Pacific. *Geophysical Research Letters* **47**, e2020GL087848, doi:<https://doi.org/10.1029/2020GL087848> (2020).
- 55 Tsui, E. Y. L. & Toumi, R. Pacific subsurface temperature as a long-range indicator of El Nino, regional precipitation, and fire. *Quarterly Journal of the Royal Meteorological Society* **148**, 2102-2117, doi:<https://doi.org/10.1002/qj.4290> (2022).
- 56 Zhang, Z.-X., Wang, J.-N. & Wang, F. A new subsurface precursor across the spring predictability barrier for the ENSO prediction. *Deep-Sea Research Part I-Oceanographic Research Papers* **203**, 104213, doi:<https://doi.org/10.1016/j.dsr.2023.104213> (2024).
- 57 Liu, C. *et al.* Subsurface ocean turbulent mixing enhances central Pacific ENSO. *Nature Communications* **16**, 2315, doi:10.1038/s41467-025-57058-4 (2025).
- 58 Zhou, L. & Zhang, R.-H. A self-attention-based neural network for three-dimensional multivariate modeling and its skillful ENSO predictions. *Science Advances* **9**, eadf2827, doi:10.1126/sciadv.adf2827 (2023).
- 59 Weisheimer, A. *et al.* Variability of ENSO Forecast Skill in 2-Year Global Reforecasts Over the 20th Century. *Geophysical Research Letters* **49**, e2022GL097885, doi:<https://doi.org/10.1029/2022GL097885> (2022).
- 60 Ehsan, M. A., L'Heureux, M. L., Tippett, M. K., Robertson, A. W. & Turmelle, J. Real-time ENSO forecast skill evaluated over the last two decades, with focus on the onset of ENSO events. *npj Climate and Atmospheric Science* **7**, 301, doi:<https://doi.org/10.1038/s41612-024-00845-5> (2024).
- 61 Hayashi, M. & Jin, F.-F. Subsurface Nonlinear Dynamical Heating and ENSO Asymmetry. *Geophysical Research Letters* **44**, 12427-12435, doi:<https://doi.org/10.1002/2017GL075771> (2017).
- 62 Hu, Z.-Z., Kumar, A., Zhu, J., Peng, P. & Huang, B. On the Challenge for ENSO Cycle Prediction: An Example from NCEP Climate Forecast System, Version 2. *Journal of Climate* **32**, 183-194, doi:<https://doi.org/10.1175/JCLI-D-18-0285.1> (2019).
- 63 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? *Bulletin of the American Meteorological Society* **93**, 631-651, doi:<https://doi.org/10.1175/BAMS-D-11-00111.1> (2012).
- 64 Larson, S. M. & Kirtman, B. P. Drivers of coupled model ENSO error dynamics and the spring

- predictability barrier. *Climate Dynamics* **48**, 3631-3644, doi:<https://doi.org/10.1007/s00382-016-3290-5> (2017).
- 65 Mukhin, D., Gavrilov, A., Seleznev, A. & Buyanova, M. An Atmospheric Signal Lowering the Spring Predictability Barrier in Statistical ENSO Forecasts. *Geophysical Research Letters* **48**, e2020GL091287, doi:<https://doi.org/10.1029/2020GL091287> (2021).
- 66 Chen, H.-C., Tseng, Y.-H., Hu, Z.-Z. & Ding, R. Enhancing the ENSO Predictability beyond the Spring Barrier. *Scientific Reports* **10**, 984, doi:<https://doi.org/10.1038/s41598-020-57853-7> (2020).
- 67 Sundararajan, M., Taly, A. & Yan, Q.-Q. Axiomatic Attribution for Deep Networks. *Proceedings of the 34th International Conference on Machine Learning* **70**, 3319 – 3328, doi:<https://doi.org/10.48550/arXiv.1703.01365> (2017).
- 68 Gulev, S. K. *et al.* in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (ed V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou) 287-422 (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021).
- 69 Hegerl, G. C., Brönnimann, S., Schurer, A. & Cowan, T. The early 20th century warming: Anomalies, causes, and consequences. *Wiley interdisciplinary reviews. Climate change* **9**, e522, doi:<https://doi.org/10.1002/wcc.522> (2018).
- 70 Abraham, J. P. *et al.* A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Reviews of Geophysics* **51**, 450-483, doi:10.1002/rog.20022 (2013).
- 71 Timmermann, A. *et al.* El Nino-Southern Oscillation complexity. *Nature* **559**, 535-545, doi:<https://doi.org/10.1038/s41586-018-0252-6> (2018).
- 72 Peng, Q. *et al.* Strong 2023-2024 El Niño generated by ocean dynamics. *Nature Geoscience* **18**, 471–478, doi:<https://doi.org/10.1038/s41561-025-01700-9> (2025).
- 73 Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:11929*, doi:<https://doi.org/10.48550/arXiv.2010.11929> (2020).
- 74 Park, J.-H., Kug, J.-S., Li, T. & Behera, S. K. Predicting El Niño Beyond 1-year Lead: Effect of the Western Hemisphere Warm Pool. *Scientific Reports* **8**, 14957, doi:<https://doi.org/10.1038/s41598->

1145 018-33191-7 (2018).

1146 75 Ham, Y.-G., Kug, J.-S., Park, J.-Y. & Jin, F.-F. Sea surface temperature in the north tropical Atlantic
1147 as a trigger for El Niño/Southern Oscillation events. *Nature Geoscience* **6**, 112-116,
1148 doi:<https://doi.org/10.1038/ngeo1686> (2013).

1149 76 Izumo, T. *et al.* Influence of the state of the Indian Ocean Dipole on the following year's El Niño.
1150 *Nature Geoscience* **3**, 168-172, doi:<https://doi.org/10.1038/ngeo760> (2010).

1151 77 Rodrigues, R. R., Campos, E. J. D. & Haarsma, R. The Impact of ENSO on the South Atlantic
1152 Subtropical Dipole Mode. *Journal of Climate* **28**, 2691-2705, doi:[https://doi.org/10.1175/JCLI-D-](https://doi.org/10.1175/JCLI-D-14-00483.1)
1153 [14-00483.1](https://doi.org/10.1175/JCLI-D-14-00483.1) (2015).

1154 78 Woo, S., Park, J., Lee, J.-Y. & Kweon, I. S. CBAM: Convolutional Block Attention Module.
1155 *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14,*
1156 *2018, Proceedings, Part VII*, 3-19, doi:https://doi.org/10.1007/978-3-030-01234-2_1 (2018).

1157 79 Vaswani, A. *et al.* Attention is all you need. *Advances in Neural Information Processing Systems* **30**,
1158 5998-6008, doi:<https://doi.org/10.48550/arXiv.1706.03762> (2017).

1159 80 Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*,
1160 doi:<https://doi.org/10.48550/arXiv.1711.05101> (2017).

1161 81 Bassett, G. & Koenker, R. Regression Quantiles. *Econometrica* **46**, 33-50,
1162 doi:<https://www.jstor.org/stable/1913643> (1978).

1163 82 Pegion, K. & Kumar, A. Does An ENSO-Conditional Skill Mask Improve Seasonal Predictions?
1164 *Monthly Weather Review* **141**, 4515-4533, doi:<https://doi.org/10.1175/MWR-D-12-00317.1> (2013).

1165 83 Chai, T. & Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? –
1166 Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **7**, 1247-1250,
1167 doi:<https://doi.org/10.5194/gmd-7-1247-2014> (2014).

1168 84 Schaefer, J. T. The Critical Success Index as an Indicator of Warning Skill. *Weather and forecasting*
1169 **5**, 570-575, doi:[https://doi.org/10.1175/1520-0434\(1990\)005<0570:TCSIAA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2) (1990).

1170 85 Ibeuchi, C. C. & Richman, M. B. Deep learning with autoencoders and LSTM for ENSO
1171 forecasting. *Climate Dynamics* **62**, 5683–5697 doi:<https://doi.org/10.1007/s00382-024-07180-8>
1172 (2024).

1173 86 Mu, M., Xu, H. & Duan, W. A kind of initial errors related to “spring predictability barrier” for El
1174 Niño events in Zebiak-Cane model. *Geophysical Research Letters* **34**, L03709,

1175 doi:<https://doi.org/10.1029/2006GL027412> (2007).

1176 87 Duan, W. & Wei, C. The 'spring predictability barrier' for ENSO predictions and its possible

1177 mechanism: results from a fully coupled model. *International Journal of Climatology* **33**, 1280-1292,

1178 doi:<https://doi.org/10.1002/joc.3513> (2013).

1179

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.docx](#)