

OCEANOGRAPHY

A self-attention–based neural network for three-dimensional multivariate modeling and its skillful ENSO predictions

Lu Zhou¹ and Rong-Hua Zhang^{2,3*}

Large biases and uncertainties remain in real-time predictions of El Niño–Southern Oscillation (ENSO) using process-based dynamical models; recent advances in data-driven deep learning algorithms provide a promising mean to achieve superior skill in the tropical Pacific sea surface temperature (SST) modeling. Here, a specific self-attention–based neural network model is developed for ENSO predictions based on the much sought-after Transformer model, named 3D-Geoformer, which is used to predict three-dimensional (3D) upper-ocean temperature anomalies and wind stress anomalies. This purely data-driven and time-space attention-enhanced model achieves surprisingly high correlation skills for Niño 3.4 SST anomaly predictions made 18 months in advance and initiated beginning in boreal spring. Further, sensitivity experiments demonstrate that the 3D-Geoformer model can depict the evolution of upper-ocean temperature and the coupled ocean-atmosphere dynamics following the Bjerknes feedback mechanism during ENSO cycles. Such successful realizations of the self-attention–based model in ENSO predictions indicate its great potential for multidimensional spatiotemporal modeling in geoscience.

INTRODUCTION

Skillful predictions for real-time ocean-atmosphere conditions remain a long-standing challenge in climate research and have lucrative scientific and economic benefits. For example, El Niño–Southern Oscillation (ENSO) is the most dominant ocean-atmosphere phenomenon in the tropical Pacific that occurs on an interannual scale, which is manifested by the basin-wide anomalies of sea surface temperature (SST) and synergetic atmospheric circulation anomalies in the climate system (1–3). A high number of studies have demonstrated its profound influence on the climate system and society through atmospheric and oceanic teleconnections (4, 5). For instance, the occurrence of ENSO induces extreme weather events around the globe and even affects global crop yields in the coming year (6, 7). Therefore, tremendous efforts have been made to understand and predict ENSO over the past several decades (8–10).

With notable advances in observations and process understanding of ENSO, enormous progress has been made in ENSO simulation and prediction (11–13). In particular, physics-based dynamical models have been crucial tools for process understanding and phenomenon predictions. However, biases in current models stemming from inadequate representations of processes hinder realistic simulations of the climate system and long-term ENSO predictions (14, 15). At present, it is still a great challenge to accurately predict ENSO more than 1 year in advance using traditional physics-based dynamic models (10, 16–18). Fortunately, recent advances in deep learning (DL) algorithms and their innovative applications to earth

sciences provide a promising way to improve modeling of natural weather and climate phenomena (19–21).

Unlike physics-based modeling approaches, data-driven DL models use neurons to describe intrinsic physical relationships from input predictors to output predictands automatically regardless of explicit physical theories, which greatly improves the accuracy of nonlinear system modeling, including ENSO prediction. For example, the skillful Niño 3.4 SST index [i.e., the area-averaged SST anomaly in the Niño region (170°W to 120°W, 5°S to 5°N)] prediction at lead times of more than 15 months has been achieved using DL models, with the spring predictability barrier (SPB) being substantially alleviated (22–24). In addition, more comprehensive prediction tasks for ENSO-related SST and simultaneous atmospheric anomaly predictions have also been made using DL models (25–27), which generate more reasonable and robust results by considering spatiotemporal dependencies in multiple anomaly fields. Note that these previous predictions are often made in terms of a simple Niño index or a single variable (e.g., SST), which apparently cannot satisfy the need for representing the spatiotemporal evolution of ENSO and understanding its mechanism.

To date, few studies have been conducted using DL models on high-dimensional multivariate modeling in geoscience owing to their limitations in traditional algorithms (28). Conventionally, for example, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are generally used as the de facto standard for spatiotemporal predictions due to their ability to impose spatial and temporal inductive bias reductions. However, some inherent properties of CNNs and RNNs hinder their applications in high-dimensional multivariate modeling. On one hand, as the resolution increases for the input fields, CNN models need to be stacked in many layers to capture the spatial teleconnections, making the model overly complex, and it consumes a lot of computing resources. On the other hand, the nonparallelizability nature of RNNs and the vanishing gradient problem make it a challenge to learn long-range series dependencies from input data. As DL techniques

Copyright © 2023 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

¹Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, and Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China; and University of Chinese Academy of Sciences, Beijing 10029, China.

²School of Marine Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China. ³Laoshan Laboratory, Qingdao 266237, China.

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

advance, more in-depth innovations for three-dimensional (3D) multivariate predictions of ENSO are desirable to fully illustrate the ocean-atmosphere dynamics underpinning ENSO evolution and enhance model interpretability.

Recently, a self-attention–based architecture named Transformer (29) and its variants have become the standard models in natural language processing (30) and image classification tasks (31) due to their scalability, efficiency, and powerful migration ability. The Transformer architecture eliminates convolution and recurrence operations and replaces them with a self-attention mechanism to establish multivariable relationships in parallel regardless of their spatial and temporal distances. Now, this DL technique has not been widely used in geoscience despite its considerable potential for high-dimensional climate system modeling (32).

In this study, we present an innovative application of a variant model based on the Transformer, named 3D-Geoformer, for spatio-temporal multivariate predictions of ENSO, with the spatiotemporal self-attention mechanism interpreted (see Materials and Methods). Further prediction examples are made to demonstrate its superior and robust modeling skills; this technique is particularly suitable for capturing the nonlocal long-range connections among atmospheric and oceanic anomalies during ENSO evolution. Specifically, three essential oceanic and atmospheric variables are selected to train the 3D-Geoformer model, including sea surface zonal (τ_x) and meridional (τ_y) wind stress, and seven-layer ocean temperature anomalies in the upper 150 m (at 5, 20, 40, 60, 90, 120, and 150 m). These variables consist of physical interactions that participate in the form of the Bjerknes feedback, providing theoretical interpretations for the realization in model predictions.

RESULTS

Our 3D-Geoformer model uses sea surface wind stress (τ_x and τ_y) and upper-ocean temperature anomaly fields from Coupled Model Intercomparison Project phase 6 (CMIP6) data over the region of (92°E to 330°E, 20°S to 20°N) for 12 consecutive months as predictors and the same anomaly fields for the following 20 months as predictands. Then, the multivariate prediction skills of the 3D-Geoformer model are comprehensively assessed with data from the Global Ocean Data Assimilation System (GODAS) reanalysis during 1980–2021 in terms of multiple spatiotemporal evolving characteristics using multiple error statistical analyses. It is noted that observed data are not used to fine-tune the 3D-Geoformer model in this study, because the transfer training technique is difficult to ensure a consistent and global optimization for such high-dimensional predictands (see the Supplementary Materials for more discussions).

Assessments of general prediction skills for the 3D-Geoformer model

To quantitatively assess the 3D-Geoformer prediction skills, we calculate multivariable correlations for the spatial distribution and Niño indices, as shown in Figs. 1 and 2. First, the anomaly correlation coefficient (ACC) distributions of τ_x , τ_y , SST, and equatorial temperature anomalies in the zonal-depth sections reveal the spatial differences in their prediction skills at different lead times. Overall, the highest ACC skill does not occur in the eastern Pacific but is located in the central basin, where large predictable signals exist with much weaker noise (33). As shown, the

predictions of upper-ocean temperature are more skillful and robust than those of wind stress. For example, the ACC skill of upper-ocean temperature fields remains larger than 0.5 over most of the tropical Pacific for 12-month lead time predictions, while the wind stress anomalies almost lose their predictability at lead times greater than 9 months.

Notably, there is a low prediction skill region for upper-ocean temperature anomalies in the equatorial western Pacific (Fig. 1, C, G, and K), which tends to extend from the sea surface to the subsurface in the central Pacific (Fig. 1, D, H, and L). This low skill pattern is also represented in the CMIP6 simulations, which can be attributed to the excessive westward extension of the equatorial Pacific cold tongue in climate models [(33–35); it can be further demonstrated by comparing the dominant patterns of observed temperature anomalies (as represented by the leading empirical orthogonal function modes) with those in CMIP6 simulations as displayed in figs. S1 and S2]. Hence, when training the 3D-Geoformer model using the biased CMIP6 data, these systematic biases are also incorrectly learned in our data-driven model, causing the predicted temperature anomalies to extend too far west of the dateline compared to the observations. This prediction bias in the 3D-Geoformer model can be reduced to some extent by implementing the transfer learning strategy during the training process (fig. S3).

To further highlight the reliable ENSO prediction skills, we focus on the performance in terms of Niño index predictions, including the Niño 3.4, Niño 3, and Niño 4 indices [i.e., area-averaged SST anomaly in the Niño 3.4 region, Niño 3 region (150°W to 90°W, 5°S to 5°N), and Niño 4 region (160°E to 150°W, 5°S to 5°N), respectively]. These indices are commonly used to define and categorize ENSO events, such as the different types of ENSO events and their specific features. Accurate predictions of Niño indices are conducive to increasing our ability to predict the intensities and different types of ENSO events. Here, we use three measures to evaluate the Niño index prediction skills relative to observations: Pearson correlation coefficient (PCC), root mean square error (RMSE), and mean absolute error (MAE), respectively.

The all-season correlation skills of the Niño indices assessed during 1983–2021 in Figs. 2 (A to C) indicate that the ENSO prediction performances of the 3D-Geoformer model outperforms the current state-of-the-art climate models (see figs. S4 to S7 for more comparison analyses). Specifically, it can create a valid Niño 3.4 index prediction for a lead time of up to 18 months (i.e., PCC > 0.5); the valid prediction lead time for both the Niño 3 and Niño 4 indices is more than 16 months, while the prediction errors (as indicated by RMSE and MAE) are acceptable. The corresponding seasonal skill analyses of the Niño index predictions are further verified (Fig. 2, D to F), indicating the superior prediction skills in the 3D-Geoformer model. Here, a noteworthy feature is that the ENSO prediction skill is particularly enhanced to more than 16 months for predictions made through the boreal springtime (i.e., PCC > 0.5; figs. S5 and S6), which is the most challenging season to predict. These quantitative analyses fully demonstrate the impressive realization using the spatiotemporal self-attention–based model to make 3D multivariate evolution predictions of ENSO for long lead time. Moreover, detailed comparisons of prediction skills from this 3D-Geoformer model are made with various other model and products in figs. S4 to S7, including the CNN model developed by Ham *et al.* (22), the Climate Prediction Center/National Oceanic and Atmospheric Administration

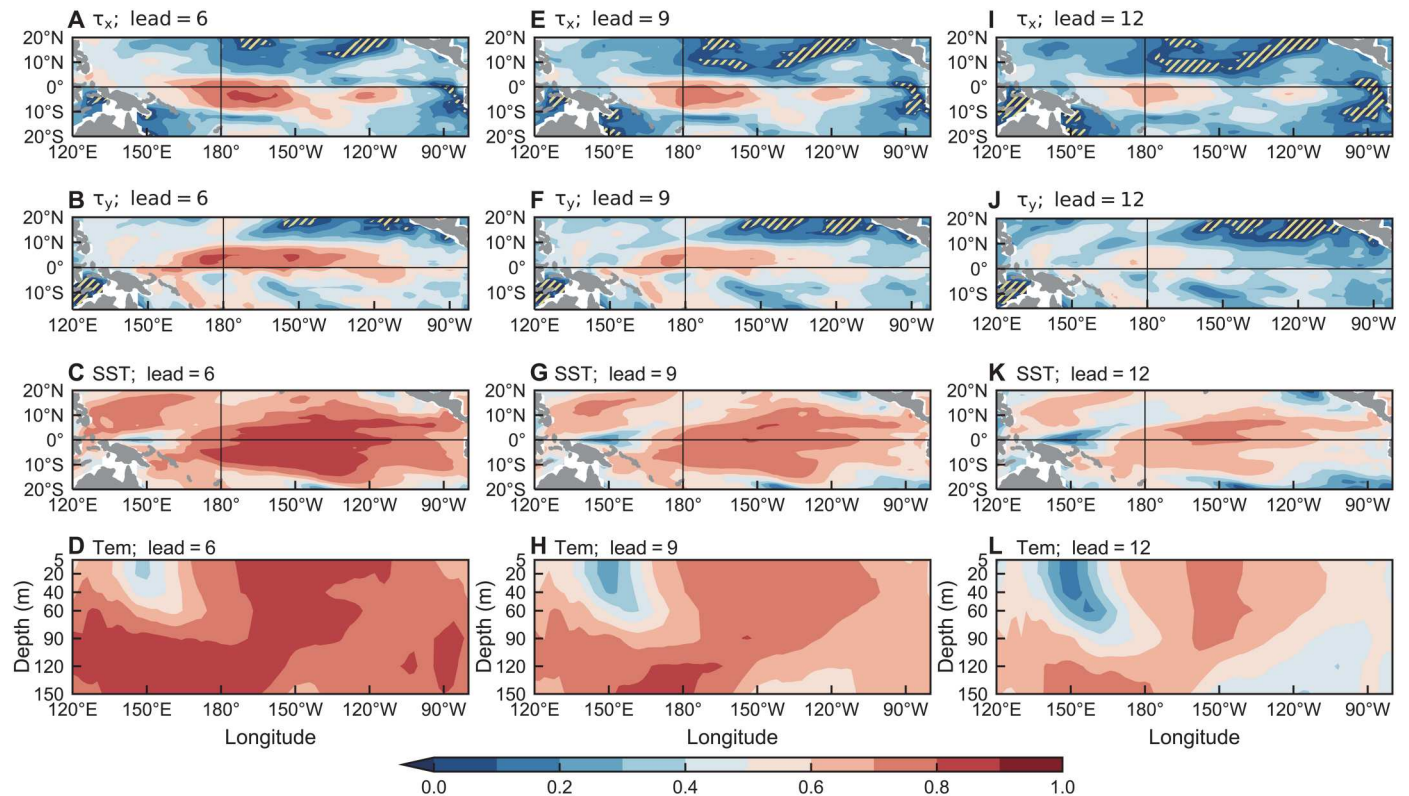


Fig. 1. Distributions of ACCs calculated between analyzed and predicted fields during 1983–2021. (A to L) The results assessed at 6-, 9-, and 12-month lead times (the first, second, and third columns) are displayed: τ_x , τ_y , and SST in the horizontal sections (the first three rows) and upper-ocean temperature anomalies in the zonal-depth sections on the equator (the last row), respectively. The shading indicates the region where ACCs are below 95% confidence level based on Student's *t* test.

(NOAA) and International Research Institute for Climate and Society prediction results (36), and the North American Multi-Model Ensemble hindcast products (37).

Skillful multivariate predictions: A case analysis for the 2015–2016 El Niño

To elaborate the covarying dynamics more clearly, we select the 2015–2016 El Niño event as an example to demonstrate the superiority of the 3D-Geoformer model in terms of 3D upper-ocean temperature predictions, and more quantitative statistical results are also presented in this subsection.

The 2015–2016 El Niño is one of the strongest ENSO events ever recorded, and large uncertainties still exist in predicting its onset when using traditional dynamical models to start predictions from spring 2015 as an initial condition (38). It is a suitable example to evaluate the effectiveness of the 3D-Geoformer model for predicting this extreme ENSO event. The model can accurately depict the upper-ocean temperature evolution and its interactions with SST and wind stress in the tropical Pacific Ocean.

Examples for the predicted Niño 3.4 SST anomalies initiated from February to June 2015 are shown in Fig. 3A, together with the corresponding observed anomalies. In general, the temporal evolution of the predicted Niño 3.4 index closely resembles the observations, despite a weaker amplitude. For example, the predictions initiated in February and March generate moderate Niño 3.4 SST anomalies of approximately 1.5°C in winter; the observed anomalies reach approximately 3°C. The underestimation prediction bias

during the springtime is due to SPB, which can be substantially ameliorated in the 3D-Geoformer model, although it cannot be eliminated completely. In contrast, starting from April, predictions can accurately capture the strong SST anomalies similar to observations.

Furthermore, we highlight the 3D upper-ocean temperature anomaly evolution in association with the surface wind stress that is captured in the 3D-Geoformer modeling (Fig. 3); as evident, ENSO prediction skills are enhanced with the multivariate synergic ocean-atmosphere dynamics represented: The 3D-Geoformer model can accurately depict the position and amplitude of covarying anomalies for more than 12 months in advance and distinguish among different types of central Pacific and eastern Pacific El Niño at lead times of more than 6 months (figs. S8 and S9).

These effective and accurate predictions using the 3D-Geoformer model can be partly attributed to adequate representations of the ocean-atmosphere system in a coupled manner, which is in accordance with the Bjerknes feedback mechanism. During El Niño, for example, large warm SST anomalies in the eastern equatorial Pacific and covarying positive temperature anomalies in the subsurface are accompanied by westerly wind anomalies over the central equatorial Pacific around the date line (fig. S10). As shown in Fig. 3 (B to M), for a 1-year multivariate prediction initiated in April 2015, the 3D temperature evolution and its relationship with wind stress are clearly demonstrated. For instance, warm ocean temperature condition is observed to be sustained in the central Pacific from 2014 on, which provides a preconditioning as a starting point for the 2015 El

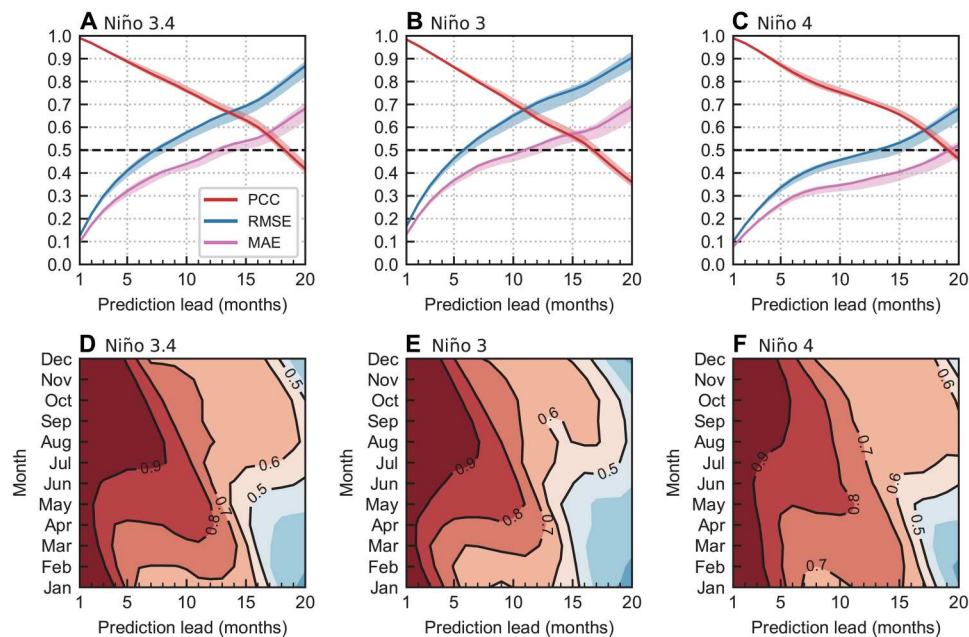


Fig. 2. The prediction skill of 3-month-averaged Niño indices assessed during 1983–2021 for the 3D-Geoformer model. (A to C) All-season PCC (red), RMSE (blue), and MAE (pink) as a function of prediction lead months for the Niño 3.4 index, Niño 3 index, and Niño 4 index, respectively. **(D to F)** The seasonality of correlation skills is further assessed as a function of lead time and calendar month for the Niño 3.4 index, Niño 3 index, and Niño 4 index, with contours of correlation skills exceeding the highlighted 0.5 value. The shading around the lines in (A) to (C) indicates the 95% confidence interval obtained by the bootstrap method; the correlation skill is shown in (D) to (F) only for that over the 95% confidence level based on Student's *t* test.

Niño event to occur, allowing for early warming up and continual strengthening; the event reaches the mature stage in December 2015 (fig. S10). Correspondingly, the 3D-Geoformer predictions reliably reproduce the observed evolution of 3D temperature anomalies and the corresponding surface wind variability: Pronounced subsurface warm temperature anomalies in the western equatorial Pacific propagate eastward along the equator from the spring of 2015, which is accompanied by persistent westerly wind anomalies over the western-central Pacific; when these subsurface anomalies arrive in the eastern basin, large positive SST anomalies are occurring in the eastern Pacific. These 3D multivariate synergies reflect the Bjerknes positive feedback mechanism well and are entirely comparable with those made by the state-of-the-art dynamic models (39).

Sensitivity experiments performed for interpretability

As noted above, the skillful realization of 3D-Geoformer modeling emphasizes the critical role of multivariate coupling dynamics as represented in the model configuration, including the synergies among wind stress, SST, and upper-ocean temperature in the tropical Pacific, as well as the effects of off-equatorial processes (1, 40, 41). In particular, the interactions among upper-ocean temperature, SST, and wind stress are implicitly characterized in our model in a monthly roll-coupled manner, enabling superior representations of their coupling. In addition, the equatorward intrusions of temperature anomalies along the Pacific North Equatorial Countercurrent pathway are also depicted in our model (fig. S11).

In this subsection, we perform sensitivity experiments to reveal the fundamental processes represented in the 3D-Geoformer. In the control experiment presented above, wind stress and upper-ocean temperature anomaly fields are all used as predictors. Then,

several sensitivity experiments are performed using testing sets to investigate how the model skill is related to different anomaly fields as predictors and their effects, which is meant to represent an attempt to improve the physical interpretability for our model. Here, four test experiments are considered for wind stress removal and one experiment for off-equatorial temperature removal; wind stress sensitivity experiments are designed to explore its effects in coupled modeling, and temperature sensitivity experiments are used to examine the contributions of off-equatorial processes to ENSO prediction. All sensitivity analyses for the prediction experiments are ensemble-averaged results from seven calculations, with only the initialization parameters used in training process being different.

Wind stress effects on ENSO predictions

Wind stress is an important component in the Bjerknes feedback for ENSO evolution. For example, it is widely accepted that westerly wind anomalies over the central-western equatorial Pacific have a substantial contribution to the development of El Niño events (8, 42, 43), which can provide key information for short-term ENSO predictions. By contrast, the wind anomalies over the eastern Pacific contribute relatively less in ENSO evolution. To understand the specific effect of wind stress on 3D-Geoformer ENSO predictions, four experiments are performed in which the effects of wind stress anomalies over the entire tropical Pacific (120°E to 80°W , 20°S to 20°N), western Pacific (120°E to 150°E , 20°S to 20°N ; τ_{WP}), central Pacific (150°E to 150°W , 20°S to 20°N ; τ_{CP}), and eastern Pacific (150°W to 80°W , 20°S to 20°N ; τ_{EP}) are purposefully removed from the input predictors, which are denoted as the $\tau_{\text{P_rem}}$, $\tau_{\text{WP_rem}}$, $\tau_{\text{CP_rem}}$, and $\tau_{\text{EP_rem}}$ experiments, respectively.

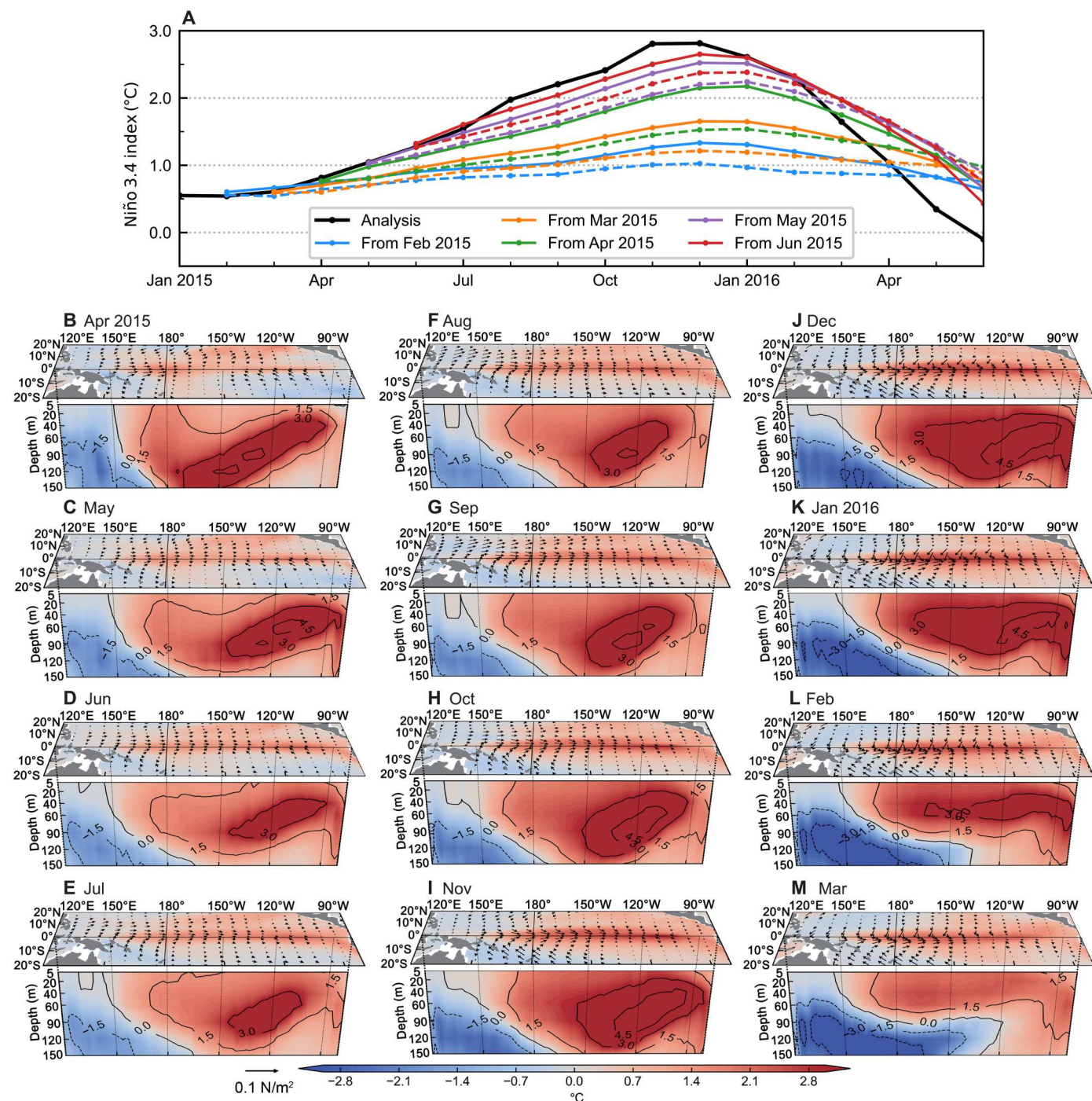


Fig. 3. Prediction examples made for the 2015–2016 super El Niño event. (A) Analyzed (black line) and 3D-Geoformer model predicted (colored lines) Niño 3.4 SST anomalies. Different colored lines indicate the predictions initiated from February to June 2015. Two types of experiments are compared with each other: Solid colored lines are for the cases in which the testing input predictors include all τ_x , τ_y , and upper-ocean temperature anomaly fields; the dotted colored lines are for the other sensitivity cases in which the testing input predictors include the upper-ocean temperature only with wind stress anomaly effects being removed. (B to M) One-year prediction example for spatiotemporal evolutions initiated from April 2015: wind stress (vectors) and SST (shading) in the horizontal sections and synergetic upper-ocean temperature anomalies in the vertical-zonal sections on the equator (shading and contours), respectively.

Here, the 2015–2016 El Niño prediction examples are made in the τ_{P_rem} experiments to demonstrate the essential role played by wind stress in the extreme event prediction (the colored dotted lines in Fig. 3A show the experimental results). Obviously, the Niño 3.4 SST prediction skill in τ_{P_rem} drops significantly compared to the control experiment, in which all wind stress and upper-ocean temperature anomalies are kept in predictors. The disparity between the two experiments is clearly magnified in the springtime-initiated predictions. For example, the τ_{P_rem} prediction made when initiated in February 2015 completely fails to predict the warming trend in the autumn and winter of 2015. Similarly, the predicted intensities of the Niño 3.4 SST anomalies initiated in March and April are also greatly underestimated compared to the observed values. These results indicate that persistent westerly wind anomalies over the western Pacific during 2014–2015 are critically important to the development of the 2015–2016 El Niño event. Thus, it is necessary to fully consider the effects of surface wind forcing to accurately predict extreme El Niño events (44, 45).

More detailed consequences of the wind stress anomaly effects over the western, central, and eastern tropical Pacific on ENSO predictions are separately examined in the τ_{WP_rem} , τ_{CP_rem} , and τ_{EP_rem} experiments. As shown in Fig. 4A, we measure the

difference ratio of the Niño 3.4 prediction skill (as indicated by PCC) in the sensitivity experiments compared with the control run, i.e., $\text{Ratio} = (\text{PCC}_{\text{sen}} - \text{PCC}_{\text{control}}) / \text{PCC}_{\text{control}}$; here, PCC_{sen} and $\text{PCC}_{\text{control}}$ represent PCC skills of Niño 3.4 predictions in the sensitivity and control experiments, respectively.

Obviously, the τ_{CP} effect makes a more important contribution to the ENSO prediction skill than τ_{WP} and τ_{EP} do at lead times less than 1 year. Notably, the τ_{EP} effect tends to worsen predictions slightly at lead times more than 6 months in the all-seasonal assessment, suggesting that the τ_{EP} effect acts as a noisy predictor component and provides little predictable information for long-term ENSO predictions in our model. These characteristics are also reflected in the seasonal variation of the PCC difference between the sensitivity and control experiments (Fig. 4, C to E), in which τ_{WP} has limited effects on ENSO predictions while the τ_{CP} effect has an obvious seasonal dependence. For example, the τ_{CP} effect greatly enhances the ENSO predictions for the boreal summer and autumn seasons (Fig. 4D), which are the target seasons with low prediction skills in the 3D-Geoformer model (Fig. 2D). In some sense, the significant alleviation of the SPB problem in the 3D-Geoformer model can be partly attributed to the effect of wind stress (especially τ_{CP}) in the modeling.

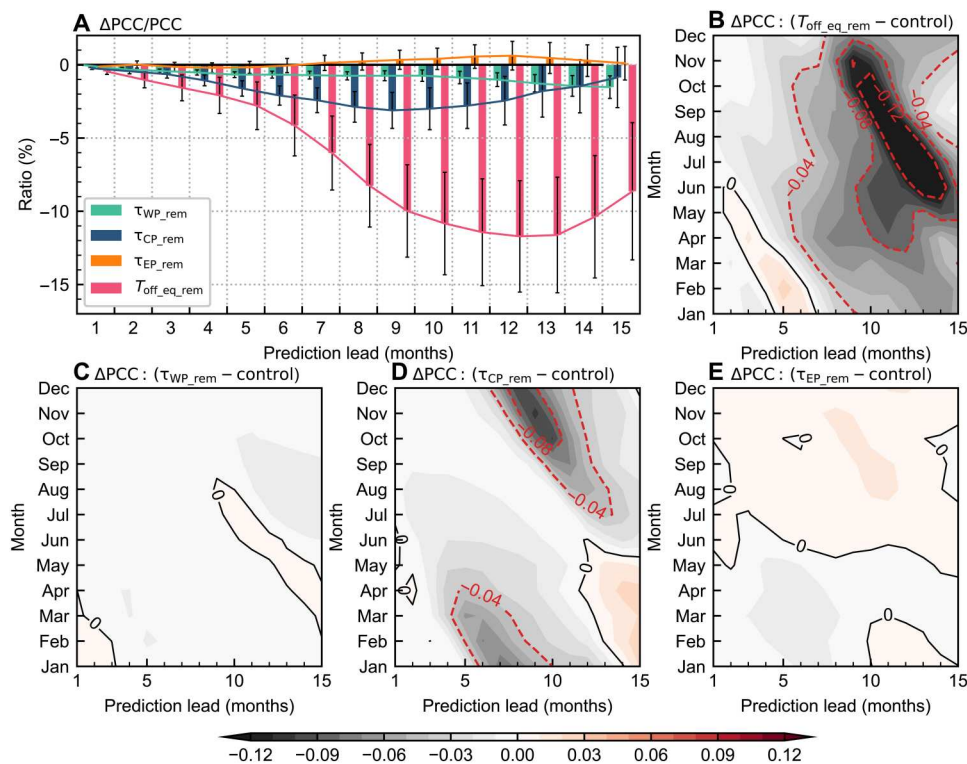


Fig. 4. Comparisons of ENSO prediction skills between the control and sensitivity experiments during 1983–2021. The effects of anomaly fields in the input predictors for the sensitivity experiments are purposely removed in some specific regions when testing the 3D-Geoformer model: the τ_{WP_rem} , τ_{CP_rem} , and τ_{EP_rem} experiments indicate that the effects of wind stress anomalies are removed from the testing predictor variables over the western (120°E to 150°E , 20°S to 20°N), central (150°E to 150°W , 20°S to 20°N), and eastern (150°W to 80°W , 20°S to 20°N) Pacific, respectively; the $T_{\text{off_eq_rem}}$ experiment indicates that the effects of upper-ocean temperature anomalies are removed off the equator (120°E to 80°W , 5°S to 20°S , and 5°N to 20°N). (A) Quantitative analyses (defined as $\Delta\text{PCC}/\text{PCC}$) of subregional anomaly effects in the τ_{WP_rem} (green), τ_{CP_rem} (blue), τ_{EP_rem} (orange), and $T_{\text{off_eq_rem}}$ (red) experiments as a function of different lead times. Here, PCC is the prediction skill in the control run and ΔPCC is the prediction skill difference between the sensitivity and control experiments. The error bar on the histogram denotes the 95% confidence interval calculated using the bootstrap method. In addition, the differences in terms of Niño 3.4 prediction PCC skills are further assessed for (B) $T_{\text{off_eq_rem}}$, (C) τ_{WP_rem} , (D) τ_{CP_rem} , and (E) τ_{EP_rem} experiments relative to the control experiment; as shown, the gray shadings indicate worse prediction skills compared with the control experiment, and the red shadings indicate better skills.

In addition to emphasizing the roles of the wind stress anomalies over the central-western Pacific in ENSO predictions, the seasonality of τ_{EP} effects merits additional investigation. Be aware that the effect of τ_{EP} on ENSO predictions (Fig. 4E) is seasonally dependent and is positive when they are initiated from springtime, especially at 6-month leads. This results are consistent with recent studies about the τ_{EP} effects on ENSO evolution (46, 47), which highlight the importance of τ_{EP} in the ENSO intensity and the temporal evolution of SST anomalies in the tropical Pacific during the decay phase of ENSO (typically occurs in boreal spring and summer). Otherwise, our sensitivity studies also demonstrate the nonnegligible role of τ_{EP} in ameliorating the SPB problem, whereas the role of τ_{EP} in ENSO predictions is currently receiving less attention. These experiments conducted on the 3D-Geoformer model reveal the potential application of data-driven DL models in process understanding.

In general, these experimental results are consistent with the physical understanding: Wind stress is an indispensable component of the ocean-atmosphere coupled system, which assures the physical representations of the dynamic principles within our model, providing short-term (less than 1 year) predictable information for ENSO predictions.

Contributions of off-equatorial temperature to ENSO predictions

As understood, long-term ENSO predictions are more dependent on subsurface temperature anomalies, including their anomaly propagation along the equatorial thermocline and the effects of off-equatorial processes. One of the advantages of the 3D-Geoformer model is its representation of upper-ocean temperature anomalies with their coherent interactions with SST and wind stress in the modeling. Here, we further demonstrate the effects of 3D upper-ocean temperature anomalies on ENSO predictions and validate the dynamic rationality of the 3D-Geoformer model.

In this section, a sensitivity experiment is performed in which the effects of off-equatorial temperature anomalies (T_{off_eq}) are removed (denoted as $T_{off_eq_rem}$) by assigning the temperature anomalies in the region (120°E to 80°W, 5°S to 20°S, and 5°N to 20°N) to zero when the temperature fields are fed into the model as predictor; then, the ENSO prediction skills are assessed relative to the control run. Figure 4 (A and B) indicates that the T_{off_eq} effect is critical in long-term ENSO predictions, which evidently improves the prediction skills when the model is initiated in the summer and autumn seasons with lead times of more than 10 months. For example, when comparing the prediction skills between the $T_{off_eq_rem}$ and control experiments, a significant difference emerges, with more than 0.12 for the predictions aimed at the July to September season.

In addition, it is noted that the $T_{off_eq_rem}$ predictions, when initiated from January to June with lead times of less than 6 months, outperform those in the control experiment (Fig. 4B), implying that the short-term ENSO predictions using the 3D-Geoformer model rely more on the τ_{CP} and equatorial temperature anomalies than those on the off-equatorial processes.

Relative effects of wind stress and 3D temperature fields

The intrinsic properties and fundamental coupling in the ocean-atmosphere system with differently time scale-dependent effects provide predictable information for ENSO predictions at different lead times. The wind stress anomaly information in predictors would contribute more to the predictions at lead times of less than 1 year, whereas subsurface temperature and off-equatorial

processes can provide interannual oceanic memory for long-term ENSO predictions (4). Then, we perform further sensitivity experiments to consistently demonstrate these inferences using the 3D-Geoformer model. Specifically, it is illustrated that τ_{CP} is more critical than τ_{WP} and τ_{EP} for predictions targeting summer and autumn seasons, and also plays a great role in mitigating the SPB problem due to stronger ocean-atmospheric interactions in the central-western Pacific. The low-frequency ocean memory contained in the T_{off_eq} effects contributes to ENSO predictions longer than 1 year in advance due to the delayed propagation of temperature anomalies along the shallow off-equatorial pathway.

All these sensitivity experiments confirm the physical rationality of inherent ocean-atmosphere coupling in the 3D-Geoformer model, providing more robust dynamic support for model prediction effectiveness. In other words, the excellent ENSO prediction performance using the 3D-Geoformer model benefits from the multivariate modeling strategy, which physically guarantees the dynamic relationships among the upper-ocean temperature, SST, and wind stress to be represented in a coupled way and thus improves the physical interpretability of the modeling success.

DISCUSSION

At present, real-time ENSO predictions mainly rely on physical-based dynamical models, and large biases and uncertainties still exist, which greatly hinder long-term ENSO predictions. Recent advances in data-driven DL-based models provide a promising way for nonlinear system modeling and multiyear ENSO predictions. However, limited by the model scale and inherent properties of traditional convolution and recurrence algorithms, accurate predictions of multidimensional ENSO evolution using DL-based models remain to be improved adequately.

Inspired by recently successful applications of the sought-after Transformer model in computer vision, we have developed a purely self-attention-based neural network, known as 3D-Geoformer, for ENSO multivariate modeling by inserting spatiotemporal attention modules into the encoder-decoder architecture. The unique properties of the self-attention mechanism in our model reinforce the long-range modeling ability to establish multivariate interconnections by performing attention analyses on the temporal and spatial axes separately. The effectiveness and superiority of the 3D-Geoformer model are demonstrated for ENSO predictions, which can depict the interactions between anomalies of 3D upper-ocean temperature and wind stress during ENSO evolution.

Specifically, the 3D-Geoformer model can predict the upper-ocean temperature anomalies in the tropical Pacific with a lead time of longer than 1 year, as represented by the ENSO-designated Niño 3.4 index prediction up to almost a year and a half. In particular, the case analysis in 2015–2016 extreme El Niño predictions fully demonstrates the impressive performance of the 3D-Geoformer model in representing the coupled ocean-atmosphere interactions. Furthermore, sensitivity experiments are conducted to corroborate the physical interpretation of our model and highlight the effectiveness of the self-attention-based modeling strategy in alleviating the SPB problem, including extreme ENSO predictions.

The black-box nature of DL-based models is a major barrier to their explainability in ENSO-related predictions. Here, we developed a 3D DL model (i.e., 3D-Geoformer model) that can capture the 3D state of ENSO, including upper-ocean temperature from the

surface to 150 m and surface winds. In terms of explainability, although most ENSO process-related results have already been reported in previous papers in a dynamical modeling context, few analyses to date have been applied to the mechanism understanding of ENSO predictions within a DL context. Here, we designed and performed sensitivity tests to nicely show where signals in predictors are important for ENSO prediction. In particular, we show how this DL method captures the critical precursors to ENSO evolution. Hence, these sensitivity experiments that are designed and demonstrated are vital to provide an advanced understanding of ENSO-related coupled dynamics and validate the dynamic soundness in a DL model. For example, the τ_{EP} experiment is designed to reveal something innovative in ENSO-related coupled dynamics of ENSO. As shown using our 3D-Geoformer, the effects of τ_{EP} on ENSO predictions are seasonally dependent and are helpful in improving spring time prediction, especially at the 6-month lead. These results are consistent with recent analyses of Peng *et al.* (46) and Zhao and Karamperidou (47), who demonstrated the non-negligible role of EP wind stress in ENSO evolution. Note that the effects of τ_{EP} on ENSO predictions and the underlying dynamics have been less studied and thus need to be better understood. Therefore, our DL-based experiments can not only offer a method for digging up coupled dynamics of ENSO but also provide an insight into some of fundamental factors affecting ENSO predictions.

There is a clear need for further improvements in some aspects of the 3D-Geoformer configuration we present in this study. On one hand, to improve computational efficiency, attention calculations in the 3D-Geoformer are performed among the fixed-size cubic patches rather than individual grid points. Hence, the effective resolution of input fields is reduced, and the attention calculations within cubic patches are lost. This calculation scheme ignores the local effects associated with vertical interactions inside one cubic patch and weakens the representation capability for multiscale processes. A more physically reasonable solution is to slice each layer of input fields separately with flexible-size windows [similar to Liu *et al.* (48)] and then calculate attention weights among these 2D patches.

On the other hand, the DL-based model performance is critically dependent on the quality and quantity of training data. The biases with the simulation data in the training set can greatly influence the model performance. A typical example from the 3D-Geoformer predictions of upper-ocean temperature is the low skill region in the equatorial western Pacific, which can be attributed to the systematic tropical bias in the CMIP6 simulations; this problem may be partially alleviated by implementing transfer training based on observational data. Although it may weaken the prediction abilities of other variables or regions when adopting our current transfer scheme (see the Supplementary Materials), we cannot deny the role of transfer learning approaches in enhancing the prediction performance of the 3D-Geoformer model. Theoretically, selecting appropriate transfer learning algorithms for specific application scenarios is still one of potential methods to improve our model's prediction skills. For example, the model-based transfer approach with shallow layer parameters frozen during the fine-tune process (49) and combined with an optimized loss function is worthy of further investigation in our DL-based modeling. At present, developing effective transfer learning approaches for such a high-

dimensional spatiotemporal prediction task is such challenging and is one of our further efforts.

In general, the successful realizations and robust performance of ENSO-related 3D multivariate predictions using the self-attention-based model indicate its conspicuous potential for high-dimensional climate modeling. The related variant networks can be directly applied to other weather and climate prediction tasks and are promising for becoming the de facto standard DL model in geoscience.

MATERIALS AND METHODS

Datasets

ENSO occurs due to its ocean-atmosphere feedback processes, which are primarily characterized by the interactions between sea surface wind stress, SST, and subsurface thermal anomalies in the tropical Pacific (1). To realistically represent these physical couplings in our study, relevant variables are adequately selected to establish the multivariate prediction model, including τ_x , τ_y , and seven-layer upper-ocean temperature anomalies.

All data need to be uniformly preprocessed before being fed into the model as an input. The detailed processing procedures are outlined as follows. First, the monthly anomaly fields are calculated by removing the long-term trend and seasonally varying climatological state. Then, the anomalies in the region (92°E to 30°W, 20°S to 20°N) are interpolated to regular grids with a resolution in the zonal direction of 2° and in the meridional direction of 0.5° (1°) (out of) 5°S to 5°N, and all land grids and missing values are assigned to a value of zero. Furthermore, the wind stress and temperature anomalies are normalized with space-averaged SD values to eliminate the influence of magnitude differences during model training. Last, all normalized fields are connected along the layer axis to construct datasets with a total of nine layers to feed into our model.

The performance of DL models is critically dependent on the quantity and quality of training data. Therefore, observational geoscience data with real-world physical processes are the optimal selection for model training, but the observational period is generally too short to meet the needs for adequate sampling. One possible option to greatly increase the quantity of training data is to use model simulation data. Thus, we also use simulation data from 23 CMIP6 climate models during 1850–2014 to train our model (table S1). In addition, the reanalysis datasets from Simple Ocean Data Assimilation (SODA) products during 1871–1979 and Ocean Reanalysis System 5 (ORAS5) products during 1958–1979 are used as validation sets to evaluate and select the model with minimal prediction biases. Last, the testing set consisting of GODAS reanalysis data during 1980–2021 is used to assess model performance and conduct sensitivity experiments.

3D-Geoformer: The multivariate prediction model and its application to geoscience

Inspired by the spatiotemporal prediction applications using the variant Transformer models (31, 50–52) and considering the unique ENSO-related ocean-atmosphere coupling, a computational efficiency spatiotemporal attention-based neural network model, known as 3D-Geoformer, has been established for 3D multivariate predictions. Similar to most competitive transduction models, the 3D-Geoformer model is also established on an encoder-decoder scheme with associated modules, including two data preprocessing

modules, encoder and decoder components, and an output layer (Fig. 5).

The 3D-Geoformer model takes the gridded data of multivariate anomalies for 12 consecutive months as input predictors (denoted as $X_{T_{in} \times C \times N_{lat} \times N_{lon}}^{in}$), consisting of τ_x , τ_y , and upper-ocean temperature anomaly fields (so the channel axis is set to $C = 9$) with a dimension of $T_{in} \times C \times N_{lat} \times N_{lon} = 12 \times 9 \times 51 \times 120$, in which $N_{lat} \times N_{lon}$ is the space grid point numbers of the original fields. The fields for the following 20 months with dimensions of $T_{out} \times C \times N_{lat} \times N_{lon} = 20 \times 9 \times 51 \times 120$ are the targeted predictands (denoted as $X_{T_{out} \times C \times N_{lat} \times N_{lon}}^{out}$).

As shown in Fig. 5A, before the encoder, the preprocessing module takes the predictors $X_{T_{in} \times C \times N_{lat} \times N_{lon}}^{in}$ as inputs. Then, each month's predictors are split into a series of fixed-size nonoverlapping patches of size $C \times h_0 \times w_0 = 9 \times 3 \times 4$ over the channel dimension and mapped to a specified dimensional symbolic representation with spatiotemporal information embedded (Fig. 5B). The symbolic representation is then fed into the following encoder, which is stacked by $n_1 = 4$ identical encoding blocks, with each block consisting of a multiheaded spatiotemporal attention layer (here, the head numbers are set to 4; Fig. 5C) and a fully connected network. The encoder component compresses the symbolic representation to a feature map matrix containing multivariate information for 12 consecutive months. Next, the encoder feature map is sent to the decoder and analyzed in the following $n_2 = 4$ decoding blocks for predictions, in which three sublayers are

contained in each block. Last, all predictions are mapped to the anomaly fields $X_{T_{out} \times C \times N_{lat} \times N_{lon}}^{out}$ with the same spatial resolutions as input predictors and the output fields for the following 20 months as targeted predictands to train the model with a rolling prediction strategy (fig. S12). That is, for each step, 1 month in the prediction, the model autoregressively generates the next prediction fields by consuming those predictions at the early prediction stages as the way in which dynamic models are used for prediction.

Model training strategy

The self-attention-based 3D-Geoformer model takes batches of variables (batch size = 8) with a length of 12 consecutive months as input predictors and the output fields for the following 20 months as targeted predictands to train the model with a rolling prediction strategy (fig. S12). That is, for each step, 1 month in the prediction, the model autoregressively generates the next prediction fields by consuming those predictions at the early prediction stages as the way in which dynamic models are used for prediction.

To quantify the multivariable fields while enhancing ENSO predictions, we combine the RMSE of multivariate fields and the Niño 3.4 index together as a loss function to measure the deviation between predictions and the target predictand

$$\text{Loss} = \frac{1}{T_{out}} \sum_{t=1}^{T_{out}} \left[\sqrt{\frac{1}{N_{lat} \times N_{lon} \times C} \sum_{i=1}^{N_{lon}} \sum_{j=1}^{N_{lat}} \sum_{k=1}^C (X_{t,k,j,i}^{out} - X_{t,k,j,i}^{tg})^2} + \sqrt{(\text{Niño}34_t^{out} - \text{Niño}34_t^{tg})^2} \right]$$

where X^{tg} is the target predictand field and Niño 3.4^{out} and Niño

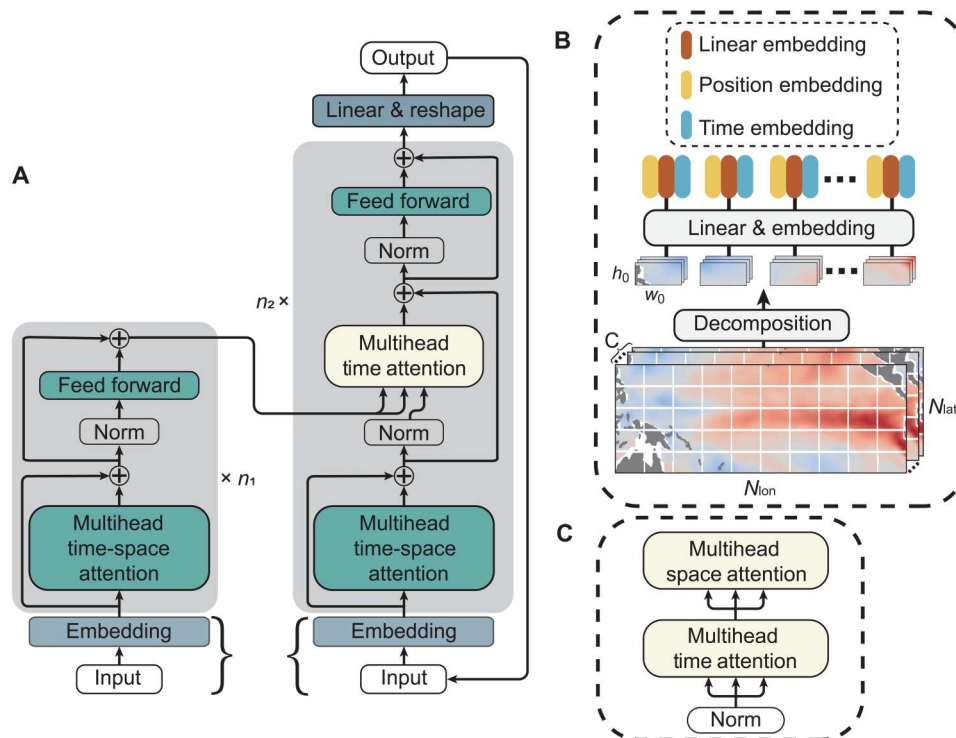


Fig. 5. Architecture of the 3D-Geoformer model for ENSO-related multivariate predictions. (A) The 3D-Geoformer model consists of two preprocessing modules at the bottoms of encoder and decoder stacks (in parentheses), an encoder-decoder structure (left and right gray boxes) based on a multiheaded spatiotemporal attention mechanism, and finally one output layer at the end of the decoder. The input predictors include τ_x and τ_y wind stress components and seven-layer upper-ocean temperature anomaly fields for 12 consecutive months in the region (92°E to 30°W, 20°S to 20°N); the anomaly fields in the following 20 months are used as the predictands for supervised training. (B) Detailed structure of the preprocessing module, which contains a field decomposition and a patch embedding process. (C) Detailed structure of the multihead spatiotemporal attention module. The data used for predictors and predictands are all normalized.

3.4⁴⁸ are the output and target Niño 3.4 SST indices, which are calculated from normalized sea temperature anomalies at 5 m depth, respectively. Then, an Adam optimization algorithm (53) is selected to optimize the 3D-Geoformer model under constraint by a learning rate warm-up trick with an initial learning rate of 1.5×10^{-4} .

After each training epoch, we evaluate the RMSE skill of the model predictions based on the validation set and save the model parameters with the minimum prediction biases. Furthermore, to reduce the prediction uncertainty in individual calculations, seven identical models are all trained in the same way but with different initialization parameters so that an ensemble prediction can be realized.

Supplementary Materials

This PDF file includes:

Supplementary Text

Figs. S1 to S12

Table S1

REFERENCES AND NOTES

1. J. Bjerknes, Atmospheric teleconnections from the equatorial Pacific. *J. Phys. Oceanogr.* **97**, 163–172 (1969).
2. D. Chen, S. E. Zebiak, A. J. Busalacchi, M. A. Cane, An improved procedure for El Niño forecasting: Implications for predictability. *Science* **269**, 1699–1702 (1995).
3. A. Timmermann, S. I. An, J. S. Kug, F. F. Jin, W. Cai, A. Capotondi, K. M. Cobb, M. Lengaigne, M. J. McPhaden, M. F. Stuecker, K. Stein, A. T. Wittenberg, K. S. Yun, T. Bayr, H. C. Chen, Y. Chikamoto, B. Dewitte, D. Dommenget, P. Grothe, E. Guilyardi, Y. G. Ham, M. Hayashi, S. Ineson, D. Kang, S. Kim, W. Kim, J. Y. Lee, T. Li, J. J. Luo, S. McGregor, Y. Planton, S. Power, H. Rashid, H. L. Ren, A. Santoso, K. Takahashi, A. Todd, G. Wang, G. Wang, R. Xie, W. H. Yang, S. W. Yeh, J. Yoon, E. Zeller, X. Zhang, El Niño–Southern oscillation complexity. *Nature* **559**, 535–545 (2018).
4. R.-H. Zhang, L. M. Rothstein, A. J. Busalacchi, Origin of upper-ocean warming and El Niño change on decadal scales in the tropical Pacific Ocean. *Nature* **391**, 879–883 (1998).
5. M. A. Alexander, I. Bladé, M. Newman, J. R. Lanzante, N.-C. Lau, J. D. Scott, The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans. *J. Climate* **15**, 2205–2231 (2002).
6. M. Luo, N.-C. Lau, Amplifying effect of ENSO on heat waves in China. *Climate Dynam.* **52**, 3277–3289 (2019).
7. W. Cai, G. Wang, B. Dewitte, L. Wu, A. Santoso, K. Takahashi, Y. Yang, A. Carreric, M. J. McPhaden, Increased variability of eastern Pacific El Niño under greenhouse warming. *Nature* **564**, 201–206 (2018).
8. M. A. Cane, S. E. Zebiak, A theory for El Niño and the Southern Oscillation. *Science* **228**, 1085–1087 (1985).
9. M. A. Cane, S. E. Zebiak, S. C. Dolan, Experimental forecasts of El Niño. *Nature* **321**, 827–832 (1986).
10. R.-H. Zhang, C. Gao, L. Feng, Recent ENSO evolution and its real-time prediction challenges. *Natl. Sci. Rev.* **9**, nwac052 (2022).
11. R.-H. Zhang, S. E. Zebiak, R. Kleeman, N. Keenlyside, A new intermediate coupled model for El Niño simulation and prediction. *Geophys. Res. Lett.* **30**, (2003).
12. R.-H. Zhang, Y. Yu, Z. Song, H.-L. Ren, Y. Tang, F. Qiao, T. Wu, C. Gao, J. Hu, F. Tian, Y. Zhu, L. Chen, H. Liu, P. Lin, F. Wu, L. Wang, A review of progress in coupled ocean–atmosphere model developments for ENSO studies in China. *J. Oceanol. Limnol.* **38**, 930–961 (2020).
13. Y. M. Tang, R.-H. Zhang, T. Liu, W. S. Duan, D. J. Yang, F. Zheng, H. L. Ren, T. Lian, C. Gao, D. K. Chen, M. Mu, Progress in ENSO prediction and predictability study. *Natl. Sci. Rev.* **5**, 826–839 (2018).
14. B. Stevens, S. Bony, Climate change. What are climate models missing? *Science* **340**, 1053–1054 (2013).
15. S. Kravtsov, C. Grimm, S. Gu, Global-scale multidecadal variability missing in state-of-the-art climate models. *NPJ Clim. Atmos. Sci.* **1**, 34 (2018).
16. M. Latif, D. Anderson, T. Barnett, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, E. Schneider, A review of the predictability and prediction of ENSO. *J. Geophys. Res. Oceans* **103**, 14375–14393 (1998).
17. E. K. Jin, J. L. Kinter, B. Wang, C. K. Park, I. S. Kang, B. P. Kirtman, J. S. Kug, A. Kumar, J. J. Luo, J. Schemm, J. Shukla, T. Yamagata, Current status of ENSO prediction skill in coupled ocean–atmosphere models. *Climate Dynam.* **31**, 647–664 (2008).
18. B. Wang, X. Luo, W. Y. Sun, Y. M. Yang, J. Liu, El Niño diversity across boreal spring predictability barrier. *Geophys. Res. Lett.* **47**, e2020GL087354 (2020).
19. C. Dong, G. Xu, G. Han, B. J. Bethel, W. Xie, S. Zhou, Recent developments in artificial intelligence in oceanography. *Ocean Land Atmos. Res.* **2022**, 2022, 1–26.
20. Y. Zhu, R.-H. Zhang, J. N. Moum, F. Wang, X. Li, D. Li, Physics-informed deep-learning parameterization of ocean vertical mixing improves climate simulations. *Natl. Sci. Rev.* **9**, nwac044 (2022).
21. G. Zheng, X. Li, R.-H. Zhang, B. Liu, Purely satellite data–driven deep learning forecast of complicated tropical instability waves. *Sci. Adv.* **6**, eaba1482 (2020).
22. Y. G. Ham, J. H. Kim, J. J. Luo, Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–572 (2019).
23. Y.-G. Ham, J.-H. Kim, E.-S. Kim, K.-W. On, Unified deep learning model for El Niño/Southern oscillation forecasts by incorporating seasonality in climate data. *Sci. Bull.* **66**, 1358–1366 (2021).
24. L. Zhou, R.-H. Zhang, A hybrid neural network model for ENSO prediction in combination with principal oscillation pattern analyses. *Adv. Atmos. Sci.* **39**, 889–902 (2022).
25. J. Taylor, M. Feng, A deep learning model for forecasting global monthly mean sea surface temperature anomalies. *Front. Clim.* **4**, 932932 (2022).
26. B. Gong, M. Langguth, Y. Ji, A. Mozaffari, S. Stadler, K. Mache, M. G. Schultz, Temperature forecasting by deep learning methods. *Geosci. Model Dev. Discuss.* **2022**, 2022, 1–35.
27. B. Mu, Y. Cui, S. Yuan, B. Qin, Simulation, precursor analysis and targeted observation sensitive area identification for two types of ENSO using ENSO-MC v1.0. *Geosci. Model Dev.* **15**, 4105–4127 (2022).
28. B. Mu, B. Qin, S. Yuan, ENSO-ASC 1.0.0: ENSO deep learning forecast model with a multivariate air–sea coupler. *Geosci. Model Dev.* **14**, 6977–6999 (2021).
29. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in *31st Conference on Neural Information Processing Systems (NIPS)*, 2017).
30. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805 [cs.CL] (11 October 2018).
31. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, An image is worth 16×16 words: Transformers for image recognition at scale. arXiv:2010.11929 [cs.CV] (22 October 2020).
32. Z. Gao, X. Shi, H. Wang, Y. Zhu, Y. Wang, M. Li, D.-Y. Yeung, Earthformer: Exploring space-time transformers for earth system forecasting. arXiv:2207.05833 [cs.LG] (12 July 2022).
33. M. Newman, P. D. Sardeshmukh, Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophys. Res. Lett.* **44**, 8520–8529 (2017).
34. G. Li, S. P. Xie, Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems. *J. Climate* **27**, 1765–1780 (2014).
35. R. Joseph, S. Nigam, ENSO evolution and teleconnections in IPCC's twentieth-century climate simulations: Realistic representation? *J. Climate* **19**, 4360–4377 (2006).
36. A. G. Barnston, M. K. Tippett, M. L. L'Heureux, S. Li, D. G. DeWitt, Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Am. Meteorol. Soc.* **93**, 631–651 (2012).
37. B. P. Kirtman, D. Min, J. M. Infanti, J. L. Kinter, D. A. Paolino, Q. Zhang, H. van den Dool, S. Saha, M. P. Mendez, E. Becker, P. Peng, P. Tripp, J. Huang, D. G. DeWitt, M. K. Tippett, A. G. Barnston, S. Li, A. Rosati, S. D. Schubert, M. Rienecker, M. Suarez, Z. E. Li, J. Marshak, Y.-K. Lim, J. Tribbia, K. Pegion, W. J. Merryfield, B. Denis, E. F. Wood, The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Am. Meteorol. Soc.* **95**, 585–601 (2014).
38. R.-H. Zhang, C. Gao, The IOCAS intermediate coupled model (IOCAS ICM) and its real-time predictions of the 2015–2016 El Niño event. *Sci. Bull.* **61**, 1061–1070 (2016).
39. M. L. L'Heureux, K. Takahashi, A. B. Watkins, A. G. Barnston, E. J. Becker, T. E. Di Liberto, F. Gamble, J. Gottschalk, M. S. Halpert, B. Huang, K. Mosquera-Vásquez, A. T. Wittenberg, Observing and predicting the 2015/16 El Niño. *Bull. Am. Meteorol. Soc.* **98**, 1363–1382 (2017).
40. R.-H. Zhang, S. Levitus, Structure and evolution of interannual variability of the tropical Pacific upper ocean temperature. *J. Geophys. Res. Oceans* **1012**, 20501–20524 (1996).
41. C. Gao, M. Chen, L. Zhou, L. Feng, R.-H. Zhang, The 2020–2021 prolonged La Niña evolution in the tropical Pacific. *Sci. China: Earth Sci.* **65**, 2248–2266 (2022).
42. S. E. Zebiak, M. A. Cane, A model El Niño–Southern Oscillation. *Mon. Weather Rev.* **115**, 2262–2278 (1987).
43. A. Capotondi, L. Ricciardulli, The influence of Pacific winds on ENSO diversity. *Sci. Rep.* **11**, 18672 (2021).
44. A. F. Z. Levine, M. J. McPhaden, How the July 2014 easterly wind burst gave the 2015–2016 El Niño a head start. *Geophys. Res. Lett.* **43**, 6503–6510 (2016).

45. A. M. Chioldi, D. E. Harrison, Observed El Niño SSTa development and the effects of easterly and westerly wind events in 2014/15. *J. Climate* **30**, 1505–1519 (2017).
46. Q. Peng, S.-P. Xie, D. Wang, Y. Kamae, H. Zhang, S. Hu, X.-T. Zheng, W. Wang, Eastern pacific wind effect on the evolution of El Niño: Implications for ENSO diversity. *J. Climate* **33**, 3197–3212 (2020).
47. S. Zhao, C. Karamperidou, Competing effects of eastern and central-western pacific winds in the evolution of the 2017 extreme coastal El Niño. *Geophys. Res. Lett.* **49**, e2022GL098859 (2022).
48. Z. Liu, Y. T. Lin, Y. Cao, H. Hu, Y. X. Wei, Z. Zhang, S. Lin, B. N. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV 2021)* (IEEE, 2021), p. 9992–10002.
49. J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.* **27**, 3320–3328 (2014).
50. M. Xu, W. Dai, C. Liu, X. Gao, W. Lin, G.-J. Qi, H. Xiong, Spatial-temporal transformer networks for traffic flow forecasting. arXiv:2001.02908 [eess.SP] (9 January 2020).
51. C. Zheng, S. J. Zhu, M. Mendieta, T. J. N. Yang, C. Chen, Z. M. Ding, 3D human pose estimation with spatial and temporal transformers, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV 2021)* (IEEE, 2021), p. 11636–11645.
52. G. Bertasius, H. Wang, L. Torresani, Is space-time attention all you need for video understanding? arXiv:2102.05095 [cs.CV] (9 February 2021).
53. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv:1412.6980 [cs.LG] (22 December 2014).

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (grant no. 42030410), Laoshan Laboratory (no. LSL202202402), the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDB40000000), and the Startup Foundation for Introducing Talent of NUIST. **Author contributions:** Conceptualization: L.Z. and R.-H.Z. Methodology: L.Z. Investigation: L.Z. and R.-H.Z. Visualization: L.Z. Supervision: R.-H.Z. Writing—original draft: L.Z. Writing—review and editing: L.Z. and R.-H.Z. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The data for this study have been deposited in the Marine Science Data Center, Chinese Academy of Science. The data (DOI: 10.12157/IOCAS.20221213.001) can be found in <http://msdc.qdio.ac.cn/data/metadata-special-detail?id=1602252663859298305>. The code (DOI: 10.5281/zenodo.7445610) can be found in <https://zenodo.org/badge/latestdoi/10.1126/sciadv.adf2827>.

Submitted 11 October 2022

Accepted 3 February 2023

Published 8 March 2023

10.1126/sciadv.adf2827

A self-attention–based neural network for three-dimensional multivariate modeling and its skillful ENSO predictions

Lu Zhou and Rong-Hua Zhang

Sci. Adv. **9** (10), eadf2827. DOI: 10.1126/sciadv.adf2827

View the article online

<https://www.science.org/doi/10.1126/sciadv.adf2827>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2023 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Supplementary Materials for
**A self-attention–based neural network for three-dimensional multivariate
modeling and its skillful ENSO predictions**

Lu Zhou *et al.*

Corresponding author: Rong-Hua Zhang, rzhang@nuist.edu.cn

Sci. Adv. **9**, eadf2827 (2023)
DOI: 10.1126/sciadv.adf2827

This PDF file includes:

Supplementary Text
Figs. S1 to S12
Table S1

Supplementary Text

1. A region with low prediction skill in the equatorial western Pacific in CMIP6 simulations and transfer learning optimizations

As shown in Fig. 1, there is region with a low prediction skill for upper-ocean temperature anomalies in the far western equatorial Pacific, which tends to extend from the sea surface vertically to the subsurface in the central Pacific. This low-skill pattern can be partly attributed to temperature bias in CMIP6 simulations used to train the 3D-Geoformer model. As a result, these systematic biases are also implicitly incorporated into the 3D-Geoformer model training. To demonstrate the spatial distribution of the simulation bias in the western Pacific, we compare the dominant temperature anomaly patterns [the leading empirical orthogonal function (EOF)] in CMIP6 simulations used in the training procedure to those in Argo observations (Figs. S1-S2; the low skill region highlighted by green box). The periods used for EOF analyses are during 2005-2021 in Argo observations and during 1985-2014 in CMIP6 historical simulations, respectively. The all leading EOFs shown in Figs. S1-S2 are normalized by multiplying the standard deviation of their corresponding principal component time series.

A notable feature of the composite dominant pattern in CMIP6 simulations is that SST anomalies during ENSO events tend to extend too far west of the dateline compared to observations (Figs. S1a-b and S2a-b); this bias pattern is widely present in individual CMIP6 simulation (the last six rows in Figs. S1-S2). Specifically, SST anomalies with opposite signs in the far western and central-eastern Pacific are seen in Argo observations, but the CMIP6 simulations typically depict an SST anomaly pattern with one anomaly sign across the entire equatorial Pacific (Fig. S1c).

In general, this prediction bias in the 3D-Geoformer model can be ameliorated by performing transfer training based on observed data. To demonstrate this, we fine-tune our models using SODA and ORAS5 reanalysis data to correct the systematic errors learned from CMIP6 samples. More details in the transfer training are as follows. First, 1500 sampling datasets are randomly selected from SODA (1850-1979) and ORAS5 (1958-1979) reanalysis products; each dataset includes predictors for twelve

consecutive months and predictands for the next twenty months with the same spatial resolution in pre-training. Then, these datasets are divided into training and validation sets according to a ratio of 9:1. The loss function used in the transfer training is identical to that in pre-training, but with a smaller learning rate of 1.5×10^{-5} . The batch size is set to 8 and the training epoch is set to 20. Prior to the transfer training, we first calculate the loss function value from the pre-trained model; then, we compare it with the subsequently fine-tuned model. The loss function values are assessed every 50 batches in the training, with smaller loss values saved for the next comparison. There are seven fine-tuned calculations based on the pre-trained models in total.

The ACC skill differences between the fine-tuned and pre-trained models are displayed in Fig. S3; the red shadings indicate the region where ACC skills get improved after the transfer training. The optimization effects due to the transfer learning differ significantly across variables and regions. For example, prediction skills for wind stress are improved more than those for sea temperature fields in most regions. Wind stress fields over the western coast of South America and the central-eastern Pacific region off the equator can be represented more accurately. In contrast, prediction skills for sea temperature anomalies are even getting worse over the most regions. The low prediction skill for upper-ocean temperature in the western Pacific region can indeed be improved from the sea surface into the subsurface to some extent (the last two rows in Fig. S3). However, since it is difficult for the transfer training to make a global optimization (actually, it can make the ENSO prediction even worse), this transfer strategy is not adopted in this paper.

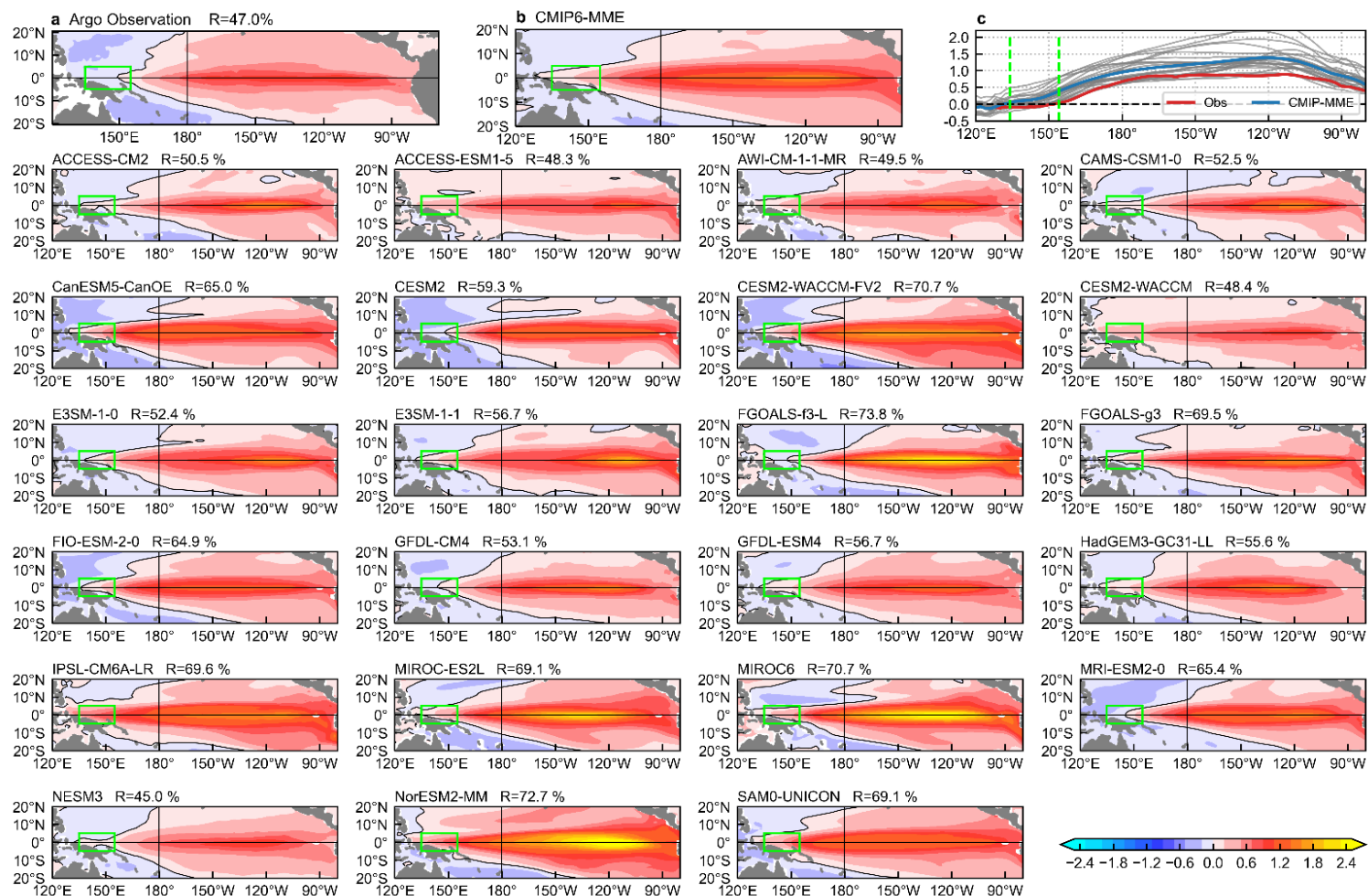


Fig. S1. (a) The dominant pattern (the leading EOF modes) of monthly SST anomalies in Argo observations during 2005-2021. (b) Same as (a) but for the ensemble mean pattern of the leading EOFs calculated from 23 CMIP6 simulations during 1985-2014. (c) Zonally averaged anomalies of the leading SST pattern along the equator (averaged between 2.5°N/S) in observations (red line), CMIP6 ensemble (blue line), and 23 CMIP6 simulations (grey lines). Other plots in the last six rows show the dominant SST anomaly pattern of each individual CMIP6 simulation; the percent variance explained by the leading EOF is also marked on the figure. The green boxes and lines highlight the region in $(135^{\circ}\text{E}-155^{\circ}\text{E}, 5^{\circ}\text{S}-5^{\circ}\text{N})$.

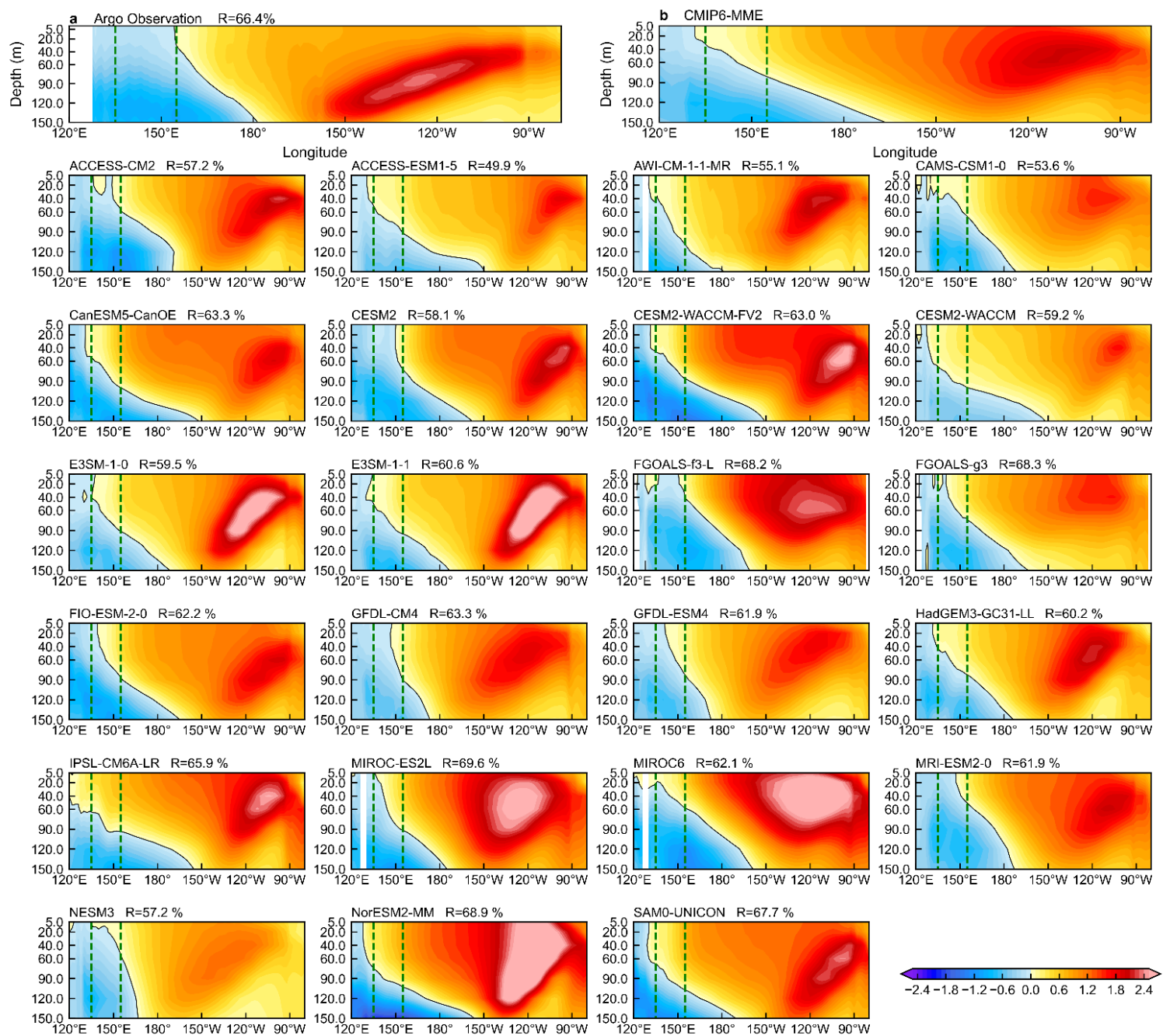


Fig. S2. Same as Fig. S1 but for the zonal-depth sections on the equator. The green lines indicate longitudes of 135°E and 155°E.

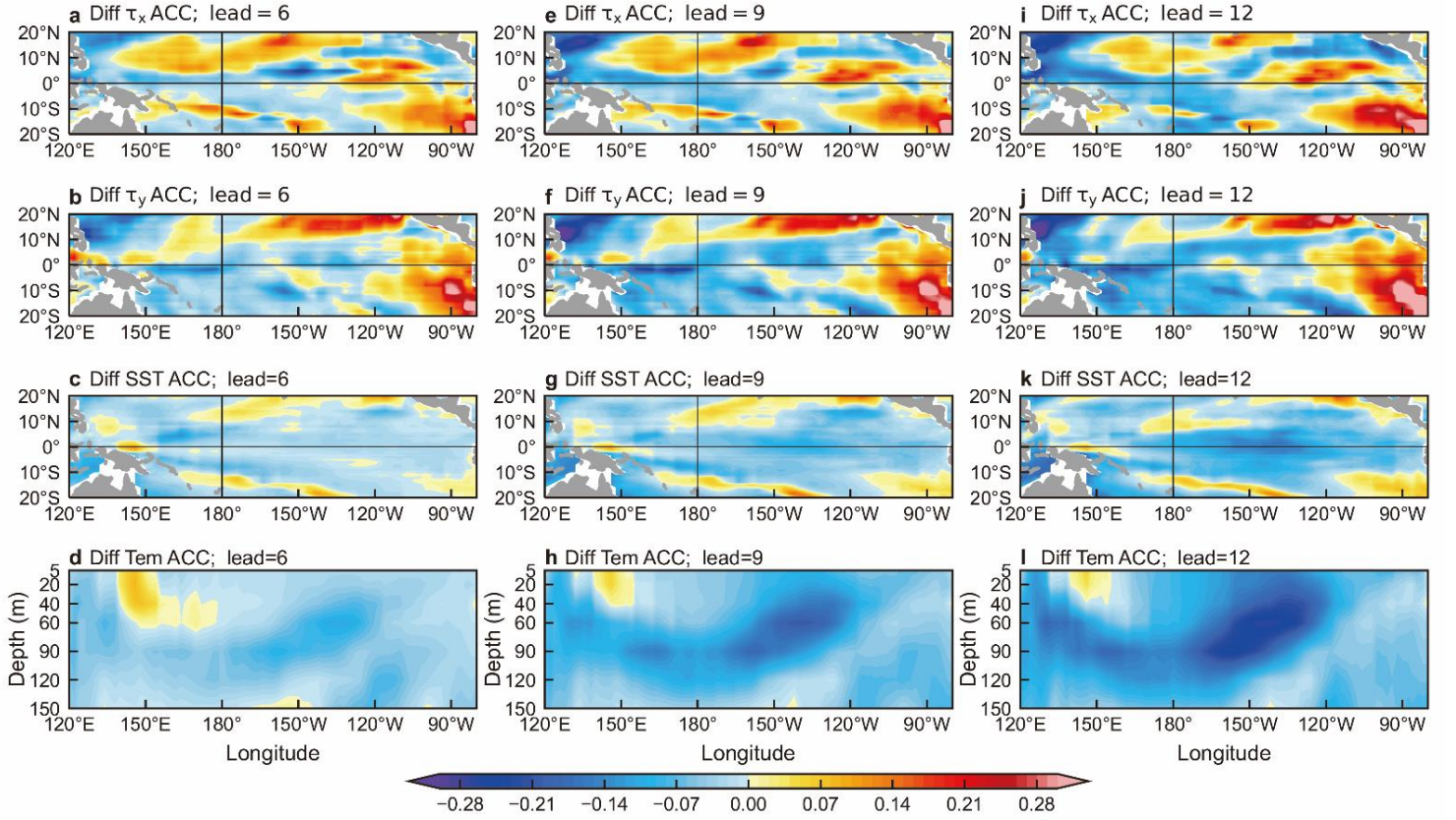


Fig. S3. Distributions of the ACC differences calculated between transfer-trained models and pre-trained models which are adopted in the article. Predictions made at 6-, 9-, and 12-month lead times (the first, second and third columns) are assessed and displayed: τ_x , τ_y and SST in the horizontal sections (the first three rows) and upper-ocean temperature anomalies in the zonal-depth sections on the equator (the last row), respectively.

2. Prediction skill comparisons with other models and products

Results from the 3D-Geoformer model are compared with those from the CNN model developed by Ham et al. (22), the Climate Prediction Center/NOAA and International Research Institute for Climate and Society (CPC/IRI) prediction results (36) and the North American Multi-Model Ensemble (NMME; 37) products. The corresponding relevant figures are shown below.

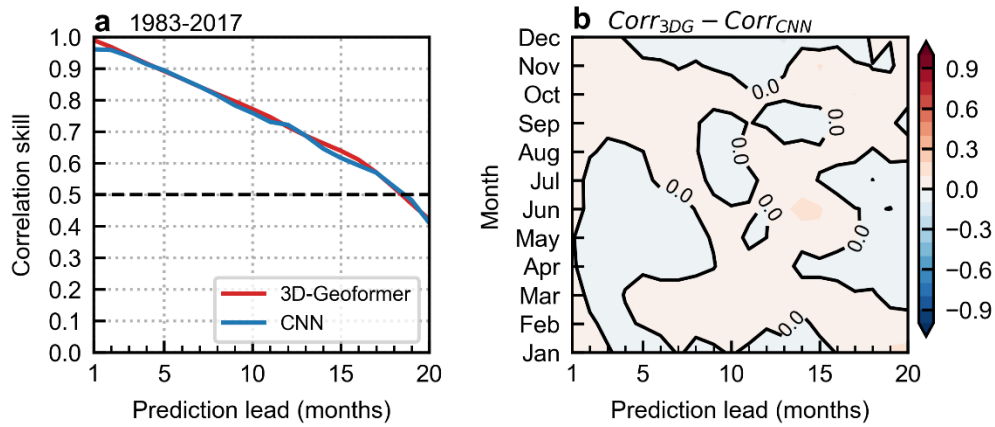


Fig. S4. Comparisons of ENSO prediction skill assessed during 1983-2017 between the 3D-Geoformer model and CNN model developed by Ham et al. (22). **(a)** The all-season PCC skill of the three-month running mean Niño 3.4 index as a function of prediction lead times for the 3D-Geoformer model (red line) and the CNN model (blue line). **(b)** The difference in the Niño 3.4 PCC calculated between the 3D-Geoformer model and CNN model as a function of the calendar month and prediction lead times.

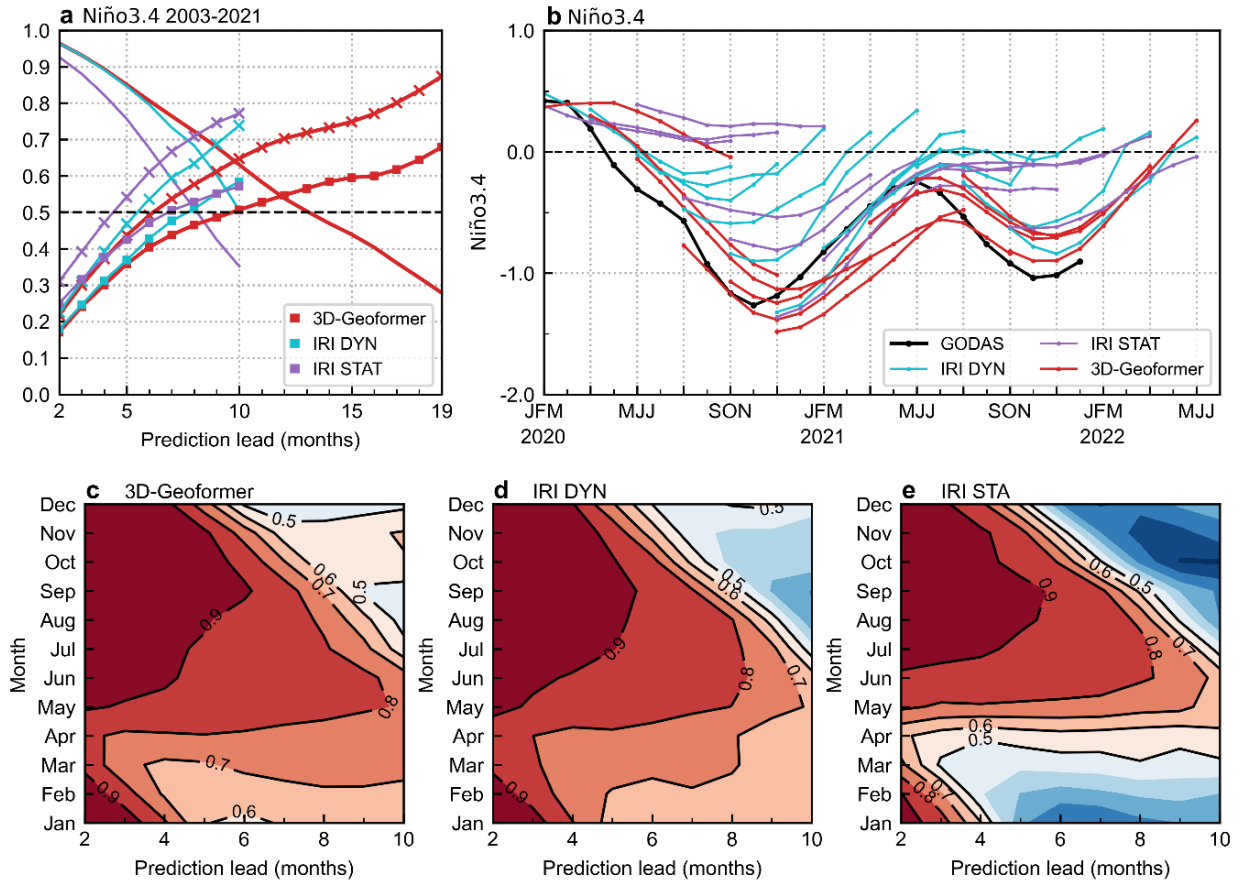


Fig. S5. Comparisons of ENSO prediction skill calculated during 2003-2021 between the 3D-Geoformer (red lines) and the multi-model ensemble mean from CPC/IRI (blue lines for dynamical model ensemble mean; purple lines for statistical model ensemble mean). **(a)** The all-season PCC (unmarked lines), RMSE (cross-marked lines), and MAE (square-marked lines) for the three-month running mean Niño 3.4 index as a function of prediction lead times. Here, the lead time is defined as the number of months between the initialized month for prediction and the middle of the three-month prediction target period. **(b)** Another example for the reanalysis (black line) and predicted (colored lines) Niño 3.4 SST anomalies during 2020-2022. Each colored line indicates the trajectory of a 9-month prediction made from initial conditions during January 2020-July 2021. Correlation skills of the Niño3.4 index as a function of lead time and calendar month in **(c)** the 3D-Geoformer predictions, **(d)** IRI dynamical ensemble-averaged predictions, and **(e)** IRI statistical ensemble-averaged predictions, with contours of correlation skills exceeding the highlighted 0.5 value.

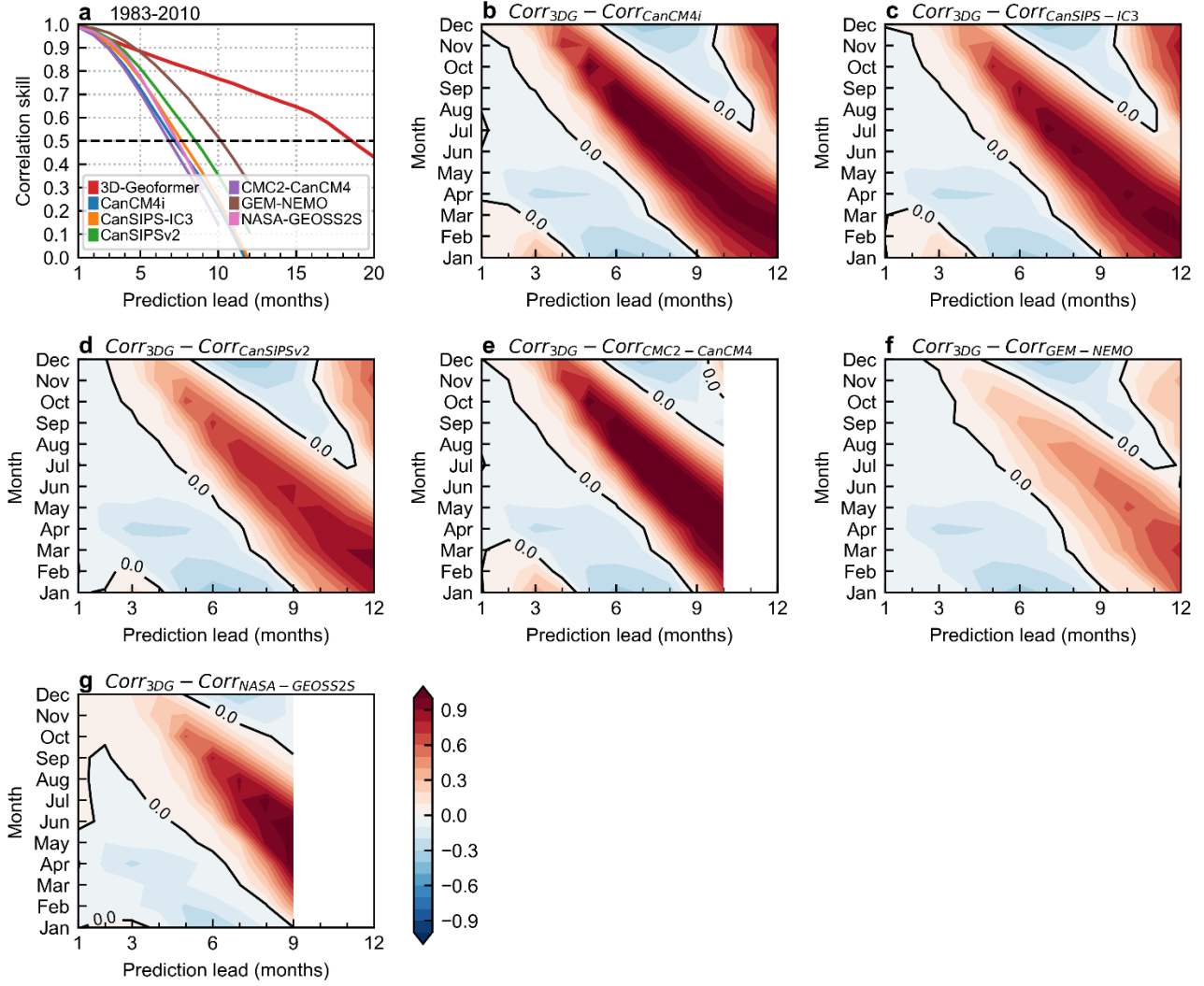


Fig. S6. Comparisons of ENSO prediction skills calculated between the 3D-Geoformer and 6 climate models from NMME products during 1983-2010. **(a)** The all-season PCC skill of the three-month running mean Niño 3.4 index as a function of prediction lead times for the 3D-Geoformer model (red line) and NMME hindcasts (other colored lines). **(b-g)** The differences in the Niño 3.4 PCC between the 3D-Geoformer model and individual models in NMME as a function of calendar month and prediction lead times.

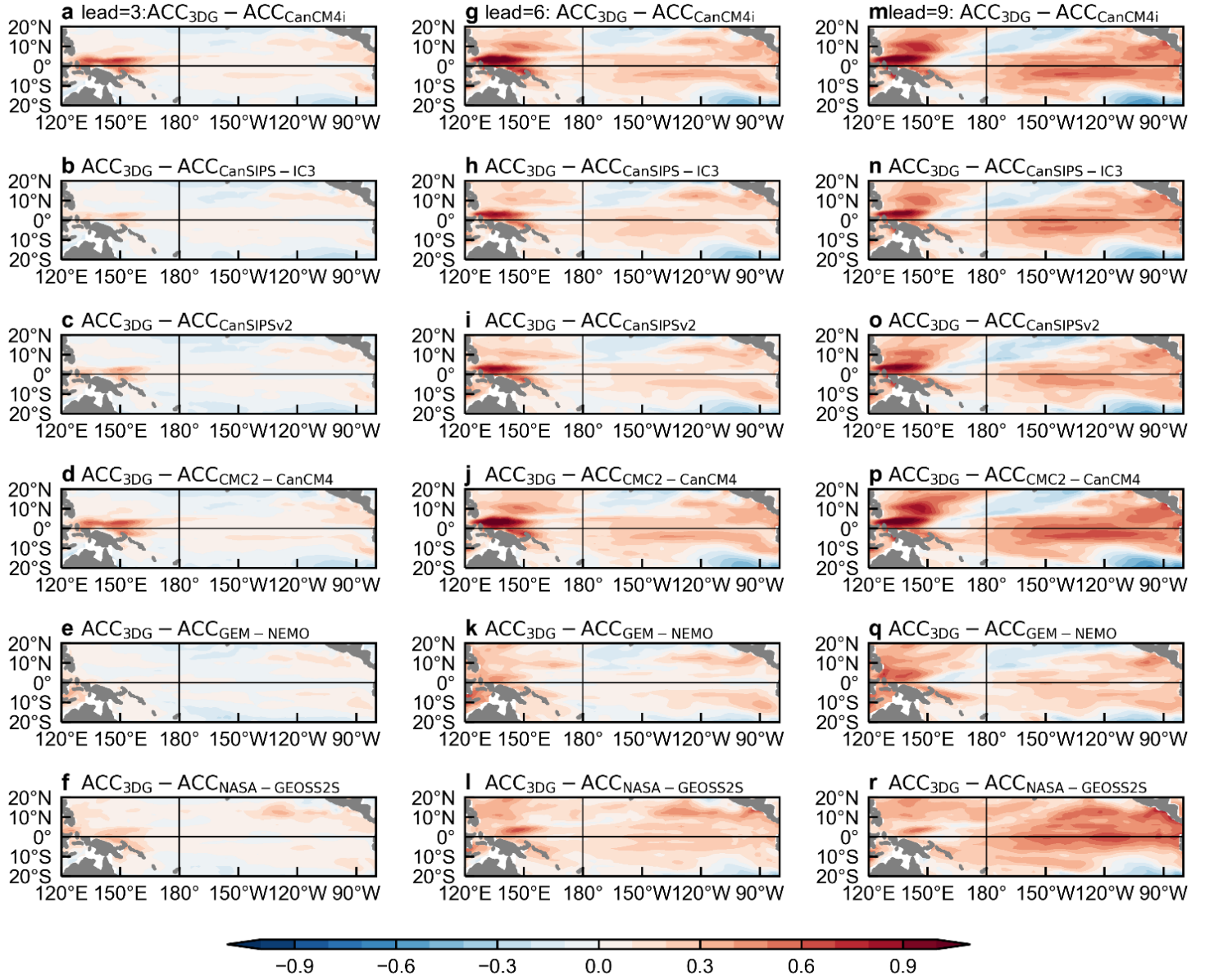


Fig. S7. Distributions of the ACC differences between the 3D-Geoformer model and climate models from the NMME products. The prediction ACC skills are assessed during 1983-2010 at 3-, 6- and 9-month lead times (the first, second and third columns), respectively.

3. Multivariate modeling and ENSO predictions

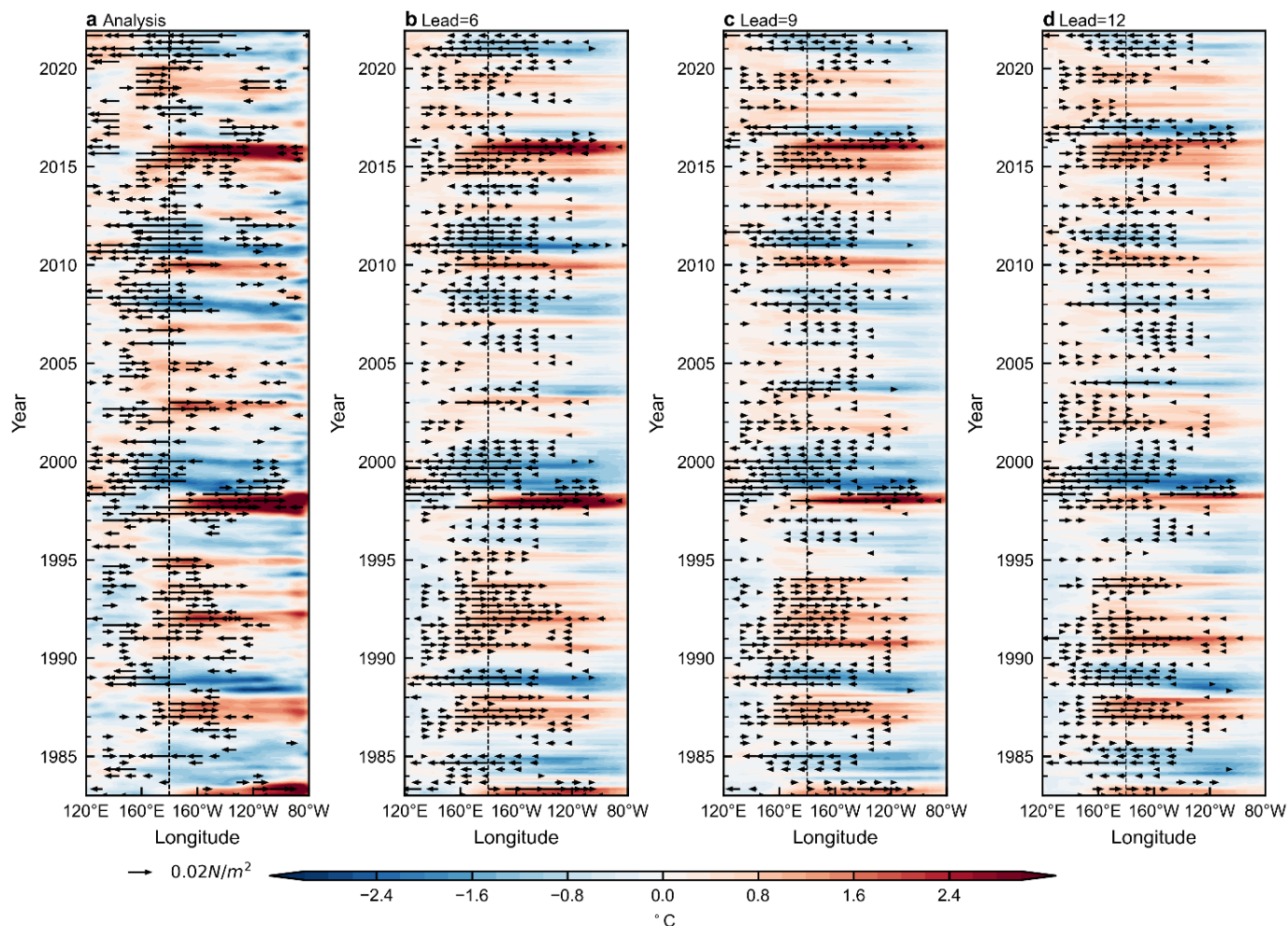


Fig. S8. Evolution of SST (shading) and τ_x anomalies (vectors) along the equator during 1983-2021. **(a)** GODAS reanalysis; **(b-d)** the 3D-Geoformer model predictions made at 6-, 9-, and 12-month lead times, respectively. The τ_x anomaly amplitude is shown only for that greater than $8 \times 10^{-3} \text{ N/m}^2$ in **(a)** and $5 \times 10^{-3} \text{ N/m}^2$ in **(b-d)**, respectively.

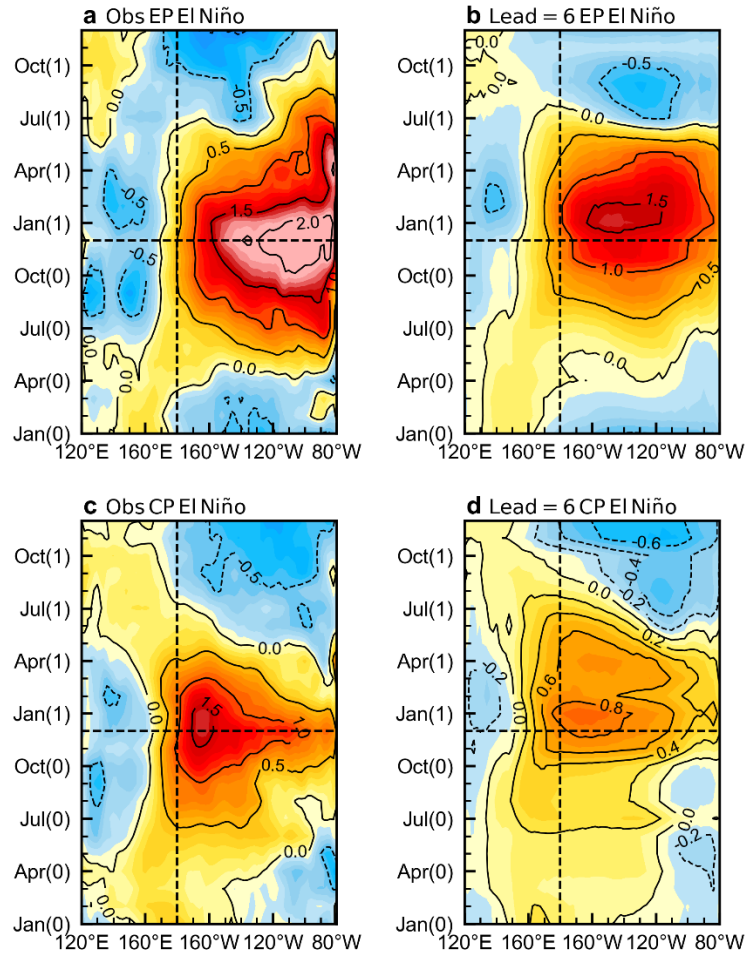


Fig. S9. Composite temporal evolutions of observed and predicted SST anomalies ($^{\circ}\text{C}$) along the equator for **(a-b)** the eastern Pacific (EP) El Niño and **(c-d)** the central Pacific (CP) El Niño during the periods from the developing year [denoted by calendar (0) on the Y-axis] to the decaying year [denoted by calendar (1) on the Y-axis]. Here, the analyses and predictions made at a lead time of 6 months are shown in the left and right columns, respectively. The EP El Niño events include 1982, 1986, 1997, and 2006; the CP El Niño events include 1991, 1994, 2002, 2004, and 2009.

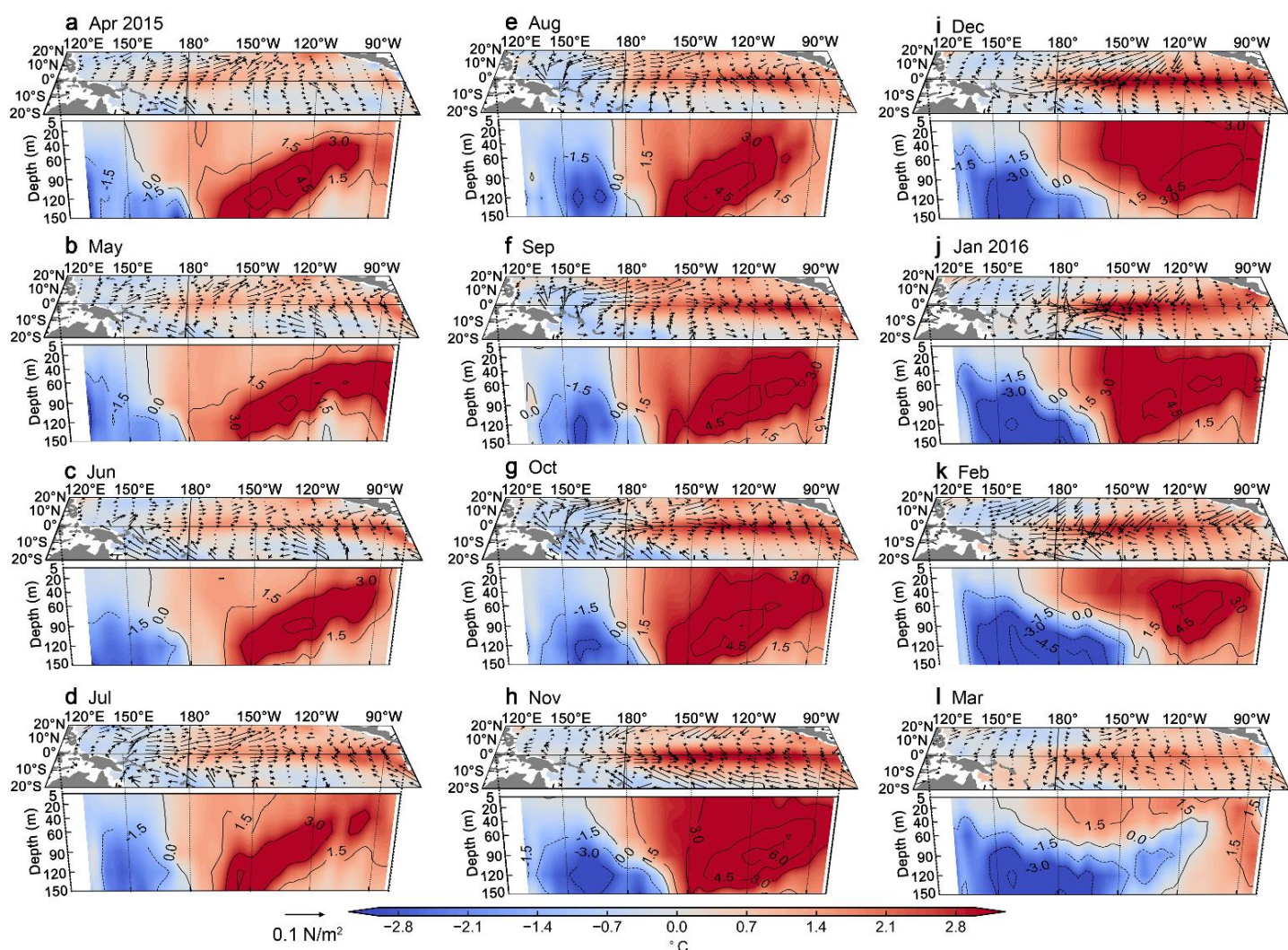


Fig. S10. Spatiotemporal evolutions shown during the periods from April 2015 to March 2016 for reanalysis wind stress (vectors) and observed SST (shading) in the horizontal sections and observed upper-ocean temperature anomalies on the equator in the vertical-zonal sections from Argo.

4. Temperature precursor propagation associated with ENSO predictions

Observations clearly indicate that ENSO evolution involves processes in the equatorial Pacific and in the off-equatorial northern Pacific near 10°N , where a well-defined shallow pathway of temperature anomalies is evident around the tropical Pacific (40). During a full ENSO cycle, the connections between the off-equatorial and equatorial regions indicate a counterclockwise pathway for temperature anomaly propagation. For example, during 1-2 years preceding the mature phase of El Niño, there are pronounced positive subsurface temperature anomalies off the equator near 10°N in the western Pacific, which tend to enter the western equatorial Pacific and then propagate eastward along the equator and arrive in the central and eastern equatorial Pacific in the following year. In some sense, these ENSO-related temperature anomalies along the pathway represent a predictable precursor for ENSO evolution. Therefore, sensitivity experiments using the 3D-Geoformer model are carried out to examine the effects of three-dimensionally propagating temperature anomalies on predictions; temperature fields as a predictor are perturbed in some selected subregions along the pathway. The experimental results are intended to further illustrate the potential for the 3D-Geoformer model to represent and learn the underlying mechanisms in a dynamic way.

Here, we divide the equatorial (5°S - 5°N) and off-equatorial (8°N - 15°N) Pacific into a series of subregions (specific ranges are labeled in Fig. S11a) and then perform sensitivity experiments with noise perturbations being added on the predictor temperature fields in each subregion; the influence on ENSO predictions is quantified (the influence degree is measured by the changes in the RMSE of the Niño 3.4 index prediction). The specific steps are outlined as follows. First, at each layer, temperature anomalies as predictors in the testing set are split into nine subregions according to the range delineated in Fig. S11a; there are seven layers in the temperature fields, so $7 \times 9 = 63$ subregions are split in total. Second, random noise of size (0, 1) is superimposed on the temperature anomaly field in the selected subregion. Third, predictions are made using the predictors with perturbed temperature fields, with the Niño 3.4 SST anomalies from the predicted temperature anomaly fields being

calculated. Finally, the saliency value $S_{i,l}$ is calculated for each subregion as follows:

$$S_{i,l} = \left| \frac{RMSE_{i,l}^{per} - RMSE_l^{control}}{Per_i \times Area_i} \right|, l=1, 2, \dots, 20; i=1, 2, \dots, 63;$$

$$RMSE_l^{control} = RMSE(Ni\tilde{no}3.4_l^{control}, Ni\tilde{no}3.4^{obs}),$$

$$RMSE_{i,l}^{per} = RMSE(Ni\tilde{no}3.4_{i,l}^{per}, Ni\tilde{no}3.4^{obs}),$$

where $RMSE_l^{control}$ and $RMSE_{i,l}^{per}$ indicate prediction skills in terms of RMSE at lead times of l months before and after the temperature perturbations in the i -th subregion, respectively; Per_i and $Area_i$ indicate the averaged perturbation noise value and the oceanic grid number in the i -th subregion. In total, twenty perturbation experiments were conducted separately for each subregion to reduce the uncertainty in the sensitivity assessments, and the results of seven 3D-Geoformer calculations were ensemble-averaged to obtain $S_{i,l}$. Briefly, the saliency value $S_{i,l}$ in each subregion reflects the dependence of ENSO predictions on different subregions at different lead times, which allows us to trace the 3D propagation pathway of temperature anomalies associated with ENSO predictions.

Figures S11**b-g** show the temperature saliency values $S_{i,l}$ in the subregions for ENSO predictions at different lead times. Overall, the ENSO predictions with lead times of less than six months are influenced strongly by the temperature anomalies in the Niño 3 and Niño 4 subregions of the upper 40 m (Fig. S11**b**). As the lead times increase, the 3D-Geoformer-based ENSO prediction skills are more dependent on temperature anomalies in the off-equatorial subregions in the central Pacific (Figs. S11**c-e**). As the ENSO predictions are made with lead times longer than one year, the prediction-affected region shifts further westward and downward to the subsurface of the tropical western Pacific (Figs. S11**f, g**). As prediction lead times increase further, the predictions show that the ENSO-related precursor (as represented by the temperature anomaly field) shifts counterclockwise along the equatorial and off-equatorial pathways. The evident propagation signal shows that the data-driven 3D-Geoformer model can represent and learn the critical dynamic mechanisms resulting

from upper-ocean temperature evolution, although there is no explicit physical constraint posed. These results clearly demonstrate the physical feasibility of the data-driven 3D-Geoformer model from a dynamic perspective, which enhances its credibility for ENSO predictions.

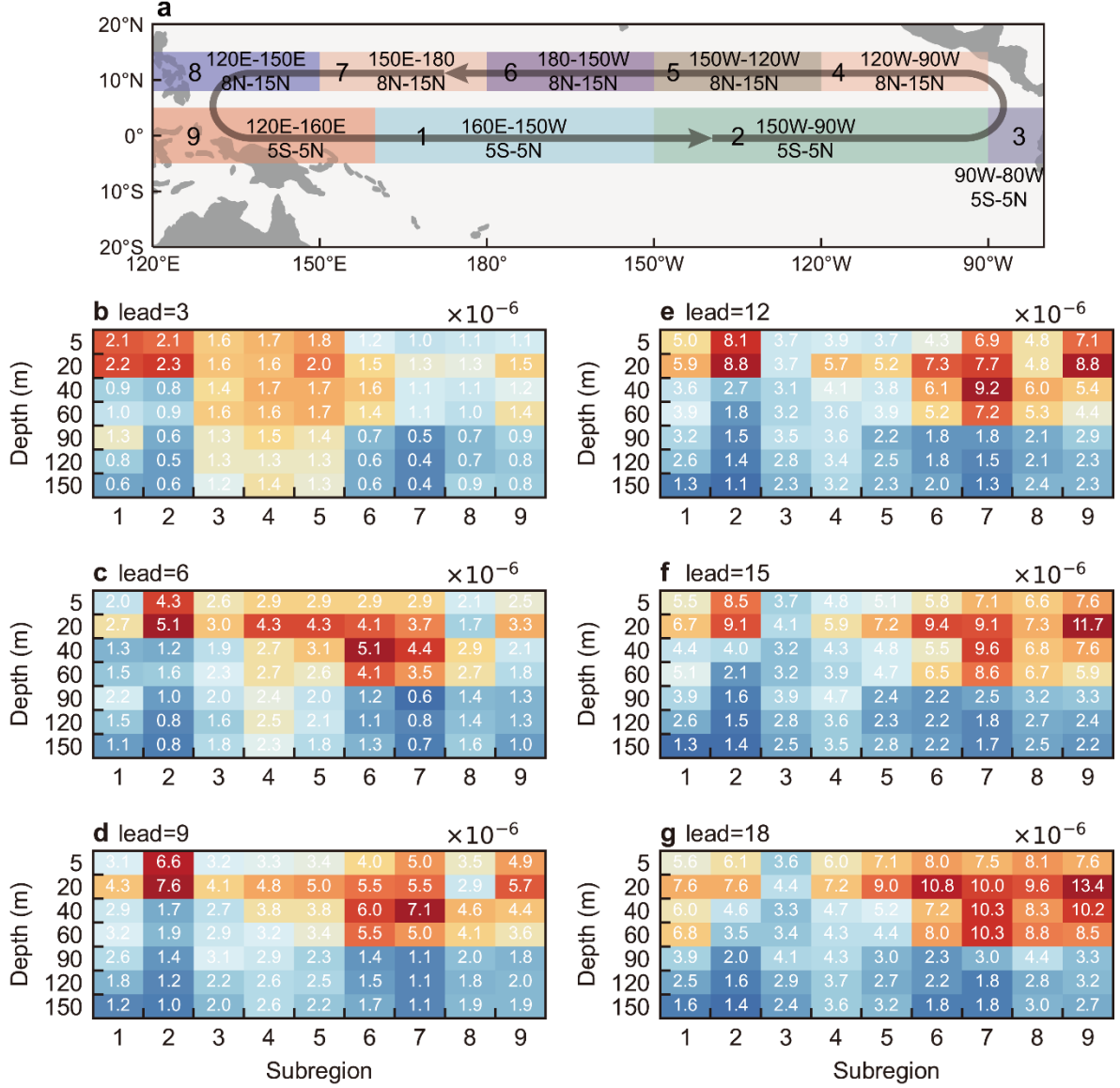


Fig. S11. Sensitivity experiments performed using the 3D-Geoformer model, in which upper-ocean temperature perturbations are made and their effects on Niño 3.4 prediction are accessed relative to the control experiment. **(a)** Distributions of the number indicating subregions into which the tropical Pacific is divided for temperature perturbation experiments performed; the specific longitude and latitude are also marked in the figure. These subregions are arranged in a counterclockwise way over the

equatorial and northern tropical Pacific to represent a cycle-like evolution of upper-ocean temperature anomalies, as indicated in Zhang and Levitus (40); the first and second subregions correspond to the Niño 4 and Niño 3 regions. **(b-g)** Sensitivity maps of ENSO predictions made using the 3D-Geoformer model for different subregions at different lead times. The X-axis indicates the divided subregion number as shown in **(a)**, and the Y-axis indicates the depth range over which the upper-ocean temperature fields are perturbed. The sensitivity results are obtained as follows. First, noise is added to predictor temperature fields in specific depth ranges and subregions; then, the predictor perturbation experiments are performed using the 3D-Geoformer; and finally, the RMSE changes in the Niño 3.4 prediction between the control and perturbation experiments is calculated to quantify the regional temperature effects on ENSO prediction. The value on each square is the absolute sensitivity value.

5. Three-dimensional Geoformer configuration and prediction procedure

The detailed 3D-Geoformer model architecture and training process are described as follows.

5.1 Preprocessing modules

There are two identical data preprocessing modules at the bottoms of the encoder and decoder stacks (Fig. 5a). The difference between the two modules is only in the sources for data receiving: the first one (at the bottom of the encoder) receives predictors as inputs, and the other receives the previously generated predictions. For simplicity, we only take the first one as an example to describe its components and data processing.

As shown in Fig. 5b, prior to the encoder, the preprocessing module takes the predictors $X_{T_{in} \times C \times N_{lat} \times N_{lon}}^{in}$ as input with dimensions of $T_{in} \times C \times N_{lat} \times N_{lon} = 12 \times 9 \times 51 \times 120$. Then, each month's predictors are split into N fixed-size nonoverlapping patches with size $C \times h_0 \times w_0 = 9 \times 3 \times 4$ over the channel dimension to obtain the decomposed predictors $X_{T_{in} \times N \times C \times h_0 \times w_0}^{in}$, in which $N = (N_{lat}/h_0) \times (N_{lon}/w_0)$ is the number of decomposed patches. Next, the decomposed predictors $X_{T_{in} \times N \times C \times h_0 \times w_0}^{in}$ are mapped to a specified dimensional embedding $X'_{T_{in} \times N \times d}$ through a learnable linear layer, where $d=256$ is the embedding dimension.

Since there are no sequential and convolutional operations in the 3D-Geoformer model, the spatiotemporal position information is not explicitly added to the analysis. Consequently, similar to Bertasius et al (52), we inject a learnable spatial positional embedding $X_{1 \times N \times d}^s$ (the yellow blocks in Fig. 5b) and a temporal embedding $X_{T_{in} \times 1 \times d}^t$ (the blue blocks in Fig. 5b) into the patch embeddings $X'_{T_{in} \times N \times d}$ to retain the spatiotemporal information for each patch. At this point, the symbolic representation $X_{T_{in} \times N \times d} = X'_{T_{in} \times N \times d} + X_{1 \times N \times d}^s + X_{T_{in} \times 1 \times d}^t$ is generated and will be fed into the following encoder.

5.2 The encoder

The encoder is a critical component for predictor compression, which maps the symbolic representation $X_{T_{in} \times N \times d}$ to a sequence of continuous feature maps relying on the spatiotemporal attention mechanism (52; Fig. 5c). Such self-attention-based operators establish spatiotemporal connections between multivariate regardless of their respective position and thus significantly reduce the dependency calculation between arbitrary positions on a constant number of operations. As shown in Fig. 5a, the encoder is stacked by $n_1 = 4$ identical encoding blocks with each block consisting of two sublayers. The first one is a multi-head spatiotemporal attention layer, and the second is a fully connected feed-forward network with a layer normalization operation. In addition, we employ a residual connection around each sublayer.

When the symbolic representation $X_{T_{in} \times N \times d}$ is delivered to the first encoding block, the spatiotemporal attention sublayer separately analyzes the dependencies in spatiotemporal dimensions via a scaled dot-product attention scoring function (29). Then, the generated feature matrix is used as the input for the next encoding block after a linear transformation. Finally, after $n_1 = 4$ rounds of the feature extraction process within the encoding blocks, the symbolic representations $X_{T_{in} \times N \times d}$ are compressed into a feature map matrix $X_{T_{in} \times N \times d}^{en}$, which contains multivariate information for twelve consecutive months and can be sent to the decoder for predictions.

5.3 The decoder and output layer

The decoder is composed of $n_2 = 4$ identical decoding blocks and three sublayers contained in each block (Fig. 5a). Similar to the encoder, the first and third sublayers in the decoder block are composed of multi-head spatiotemporal attention modules and a feed-forward network. Additionally, there is a second layer inserted in the blocks that performs spatiotemporal attention calculations between the encoder feature matrix $X_{T_{in} \times N \times d}^{en}$ and the representations processed by the first sublayer. Additionally, residual connections are applied around each sublayer in the decoder. Finally, connected with the decoder component is an output layer, which consists of a

feed-forward network to map and reshape the decoder-refined representations to the original resolutions.

The data processing of the decoder is different from that of the encoder. As shown in Fig. 5a, the decoder receives inputs from both sides, one from the bottom of the decoder stacks and the other from the second sublayer in each block, wherein the former inputs are obtained from the outputs of the preprocessing module and the latter is the feature matrix $X_{T_{in} \times N \times d}^{en}$ generated in the encoder. These inputs are analyzed in the following $n_2 = 4$ decoding blocks and output layer to generate the prediction fields $X_{T' \times C \times N_{lat} \times N_{lon}}^{out}$ having the same spatial resolutions as predictors. Here, $T'=1, 2, \dots, T_{out}$ is the time series length of the predicted fields for each step, which is extended to $T_{out}=20$ after the whole prediction (Fig. S12).

Table S1. Details of the CMIP6 simulations and reanalysis data used in this study. The CMIP6 and GODAS datasets are used for training and testing sets; SODA and ORAS5 datasets are used for verification sets.

No.	CMIP6 Model Source	Modeling Group	Period
1	ACCESS-CM2	CSIRO-ARCCSS	Jan. 1850- Dec. 2014
2	ACCESS-ESM1-5	CSIRO	
3	AWI-CM-1-1-MR	AWI	
4	CAMS-CSM1-0	CAMS	
5	CanESM5-CanOE	CCCma	
6	CESM2	NCAR	
8	CESM2-WACCM	NCAR	
7	CESM2-WACCM-FV2	NCAR	
9	E3SM-1-0	E3SM-Project LLNL UCI	
10	E3SM-1-1	E3SM-Project RUBISCO	
11	FGOALS-f3-L	CAS	
12	FGOALS-g3	CAS	
13	FIO-ESM-2-0	FIO-QLNM	
15	GFDL-CM4	NOAA-GFDL	
14	GFDL-ESM4	NOAA-GFDL	
16	HadGEM3-GC31-LL	MOHC NERC	
17	IPSL-CM6A-LR	IPSL	
19	MIROC6	MIROC	
18	MIROC-ES2L	MIROC	
20	MRI-ESM2-0	MRI	
21	NESM3	NUIST	
22	NorESM2-MM	NCC	
23	SAM0-UNICON	SNU	
No.	Reanalysis Data	Institution	Period
1	SODA	University of Maryland	Jan. 1871-Dec. 1979
2	ORAS5	ECMWF	Jan. 1958-Dec. 1979
3	GODAS	NCEP	Jan. 1980-Dec. 2021

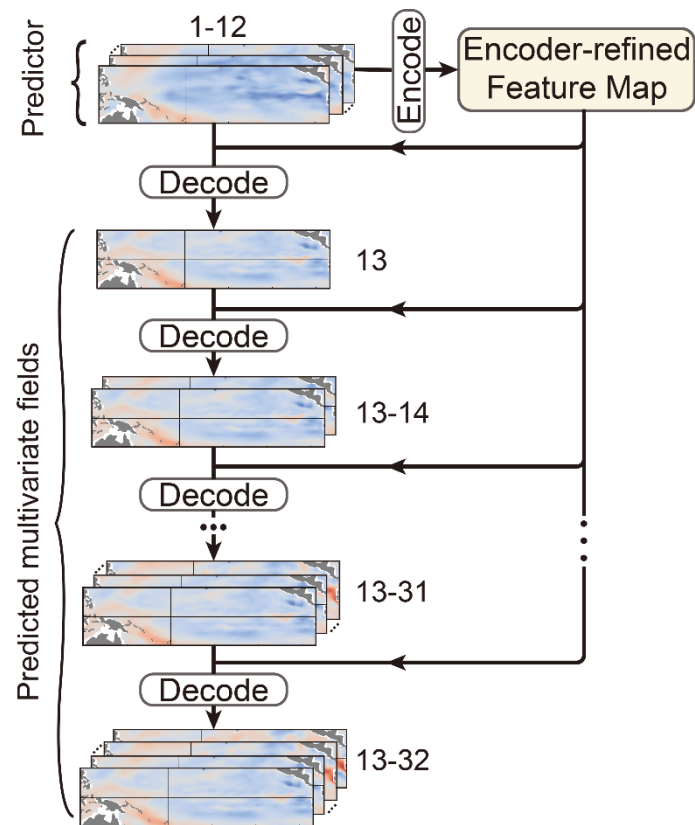


Fig. S12. Diagram showing the 3D-Geoformer model to make predictions in a rolling manner. The numbers marked in the figure indicate consequent time periods (in months; e.g., 1-12) during which anomaly fields are used as input predictor and subsequent time period (e.g., 13) at which anomaly field is predicted as targeting predictand. First, the encoder extracts spatiotemporal information from predictor fields (e.g., for 1-12 months) and compresses them into a feature map. Subsequently, the decoder consumes the previously generated prediction results and the encoder-refined feature map to again generate the next prediction fields. At each prediction step, the decoder considers all information in the previous months until all the predictions are generated for the following twenty consecutive months.