ENSO-ASC 1.0.0: ENSO Deep Learning Forecast Model with a Multivariate Air-Sea Coupler

Bin Mu[1*], Bo Qin[1*], and Shijin Yuan[1]

[1]School of Software engineering, Tongji University, Shanghai, 201804, China

5 [*]These authors contributed equally

Correspondence to: Shijin Yuan (yuanshijin2003@163.com)

Abstract. ENSO is an extremely complicated ocean-atmosphere coupling event, the development and decay of which are usually modulated by the energy interactions between multiple physical variables. In this paper, we design a multivariate air-sea coupler (ASC) based on graph using features of multiple physical variables. On the basis of this coupler, an ENSO deep

10 learning forecast model (named ENSO-ASC) is proposed, whose structure is adapted to the characteristics of the ENSO dynamics, including the encoder/decoder for capturing/restoring the multi-scale spatial-temporal correlations, and two attention ~~components~~ weights for grasping the different air-sea coupling strengths on different start calendar months and varied ~~contributions~~ effects of physical variables in ENSO amplitudes. In addition, two datasets ~~at different~~adjusted to the same resolutions are used to train the model. We firstly tune the model performance to optimal and compare it with the other state-

15 of-the-art ENSO deep learning forecast models. Then, we evaluate the ENSO forecast skill from the contributions of different predictors, the effective lead time with ~~the~~ different start calendar months, and the forecast spatial uncertainties, further analyze the underlying ENSO mechanisms. Finally, we make ENSO predictions over the validation period from 2014 to 2020. Experiment results demonstrate that ENSO-ASC outperforms the other models. Sea surface temperature (SST) and zonal wind are two crucial predictors. The correlation skill of Niño3.4 index is over 0.78/0.65/0.5 within the lead time of 6/12/18 months.

20 From two heat map analyses, we also discover the common challenges in ENSO predictability, such as the forecasting skills declining faster when making forecasts through June-July-August and the forecast errors more likely showing up in the western and central tropical Pacific Ocean in longer-term forecasts. ENSO-ASC can simulate ENSO with different strengths, and the forecasted SST and wind patterns reflect obvious Bjerknes positive feedback mechanism. These results indicate the effectiveness and superiority of our model with the multivariate air-sea coupler in predicting sophisticated ENSO and analysing

25 the underlying dynamic mechanisms.

1 Introduction

El ~~Niño~~Niño-Southern Oscillation (ENSO) can induce the global climate extremes and ecosystem impacts (Zhang et al., 2016), which is the dominant source of interannual climate change. The El ~~Niño~~Niño (La Niña) is the ocean phenomena of ENSO, and is usually considered as the large-scale positive (negative) sea surface temperature (SST) anomalies in the tropical Pacific

30 Ocean. ~~Niño~~Niño3 (~~Niño~~Niño4) index is the common indicator for ENSO research to measure the cold tongue (warm pool) variabilities, which is the averaged SST anomalies covering the ~~Niño~~Niño3 (~~Niño~~Niño4) region (See Fig. 1). Besides these two indicators, the ONI (oceanic Niño index, 3 month running mean of SST anomalies in the Niño3.4 region) has become the de-facto standard to identify the occurrence of El ~~Niño~~Niño/La Niña event: If the ONIs of 5 consecutive months are over 0.5 ℃ (below -0.5 ℃), El ~~Niño~~Niño (La Niña) occurs.
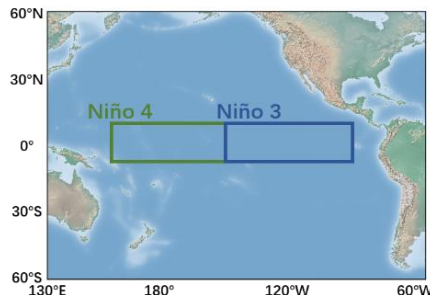

35

Figure 1: Most concerned regions in ENSO events. The blue rectangle covers the Niño3 region (~~N~~5°N-~~S~~5°S, ~~W~~150°W-~~W~~90°W), and the green rectangle covers the Niño4 region (~~N~~5°N-~~S~~5°S, ~~E~~160°E-~~W~~150°W).

Conventional forecast approaches majorly rely on numerical climate models. However, it is worth noting that the model biases of traditional approach have always been a problem for accurate ENSO predictions (Xue et al., 2013). In addition, many
40 other intrinsic factors also limit the ENSO predictability such as natural decadal variations in ENSO amplitudes. For example, predictability tends to be higher when the ENSO cycle is strong than when it is weak (Barnston et al., 2012; Balmaseda et al., 1995; McPhaden, 2012). Recently, due to deluges of multi-source real-world geoscience data starting to accumulate, e.g. remote sensing and buoy observation, meteorological researchers are inspired to build ~~more accurate~~ lightweight and convenient data-driven models at a low computational cost (Rolnick et al., 2019), which leads to a wave of formulating ENSO
45 forecast with deep learning techniques, producing more skilful ENSO predictions (Ham et al., 2019).

In the field of deep learning, ENSO prediction is usually regarded as forecasting the future evolution tendency of SST and related Niño indexes directly, subsequently ~~analyzing~~analysing the associated sophisticated mechanisms and measuring the intrinsic characteristics such as intensity and duration. Therefore, the simplest but most practical forecast ~~approaches~~ manners can be divided into two categories intuitively: Niño index forecast and SST pattern forecast.

50 As for Niño index forecast, many favourable neural networks have made accurate predictions in 6, 9 and 12 months ahead. For instance, ensemble QESN (McDermott and Wikle, 2017), BAST-RNN (McDermott and Wikle, 2019) and LSTM (long short-term memory) (Broni-Bedaiko et al., 2019) are representative works. These studies demonstrate that the deep learning can well capture the nonlinear characteristics of non-stationary time series and attain outstanding regression on Niño index.

Notwithstanding the successful attempts on the Niño index regression, there still exist many pitfalls in measuring ENSO
55 forecast skills~~s~~ ~~using~~ by only one single scalar~~.~~, For example, the important spatial-temporal energy propagation and evolution cannot be described by the indexes. ~~which~~It may lead to the blind pursuit of the accuracy of a certain indicator while seriously hampering the grasp of underlying physical mechanism. Therefore, many studies are suggestive of exploiting spatial-temporal

2

dependencies and predicting the evolution of SST patterns. Ham et al. (2019) applyies transfer learning (Yosinski et al., 2014) on historical simulations from CMIP5 (Coupled Model Intercomparison Project phase5, Bellenger et al., 2014) and reanalysis data with a CNN model to predict ENSO events, resulting in a robust and long-term forecast for up to 1.5 years, which outperforms the current numerical predictions. (Though the output of their model is still the Niño3.4 index, they construct the model and make forecasts by absorbing the historical spatial-temporal features from variable patterns instead of previous index records, so we mark this study as SST pattern forecasts in this paper) Mu et al. (2019) and He et al. (2019) both build a ConvLSTM (Shi et al., 2015) model to capture the spatial-temporal dependencies of ENSO SST patterns over multiple time horizons and obtain better predictions. Zheng et al. (2020) constructs a purely satellite data-driven deep learning model to forecast the evolutions of tropical instability wave, which is closely related to ENSO phenomena, and obtain accurate and efficient forecasts. These deep learning models tend to simulate the behaviours of numerical climate models, the inputs of which are historical geoscience data and the outputs of which are the forecasted SST pattern.

The reason for the great progress in these works is no accident. On the one hand, the deep learning model has much more complex structures and can mine the complicated features hidden in the samples more effectively, which allows them to be substantially more expressive with blending the non-stationarity in temporal and the remote multi-scale correlation teleconnections in spatial. On the other hand, it is very convenient to migrate deep learning computer vision technologies to ENSO forecast due to the nature analogy between the format of image/video frame data and meteorological time-series grid data, which offers promises for extracting spatial-temporal mechanisms of ENSO via advanced deep learning technologies relatively easily. Therefore, the data-driven deep learning can be a reliablen alternative to traditional numerical models and a powerful tool for the ENSO forecasting.

However, there are still some obstacles in the deep learning modelling process for ENSO forecast. Very often, most existing models are confined to limited or even single input predictor, such as only usinge historical SST (and wind) data as the model input. Meanwhile, the climate deep learning models are rarely adaptively customized to the specific physical mechanisms of ENSO. These situations lead to poor interpretability and low confidence of ENSO-related deep learning models. ENSO is an extremely complicated ocean-atmosphere coupling event, the development and decay phases are closely associated with some crucial dynamic mechanisms and Walker circulation (Bayr et al., 2020), whose status have great impacts. Walker circulation is usually modulated by multi-physical variables (such as SST, wind, precipitation, etc.), and there are always coupled coupling interactions between different variables. More specifically, the varieties in Walker circulation have strong temporal-lag effects on ENSO ("memory effects").; Tthe position of the ascending branch is also a very important climatic condition for the occurrence of El Niño. Such priori ENSO knowledge has not been effectively used in deep learning model.

Therefore, in order to further improve the ENSO prediction skill, there is an essential principle that should be reflected in climate deep learning models: Ssubjectively incorporating priori ENSO knowledge into the deep learning formalization and deriving hand-crafted features to make predictions.

In this paper, according to the important synergies of multiple variables in crucial ENSO dynamic mechanisms and Walker circulation, we select 6 indispensable variables (SST, u-wind, v-wind, rain, cloud, and vapor) that are induced from

ENSO-related key processes to build a multivariate air-sea coupler (ASC) based on graph mathematically, which emphasizes the energy exchange between multiple variables. We then leverage th~~i~~se coupler ~~and~~ to build up the ENSO deep learning

forecast model, named ENSO-ASC, with an encoder-coupler-decoder structure to extract the multi-scale spatial-temporal

95 features of multiple physical variables. Two attention ~~components~~ weights are also proposed to grasp the different air-sea coupling strength~~s~~ on different start calendar months and varied ~~contributions~~ effects of these variables. A loss function combining *MSE* and *MAE* is used to guide the model training precisely， *SSIM* (structural similarity) (Wang et al., 2004) and *PSNR* (peak signal to noise ratio) are used as metrics to evaluate the spatial consistency of the forecasted patterns.

Two datasets are applied for model training to ensure that the systematic ~~system~~ forecast errors are fully corrected after

100 tuning by the higher quality dataset:~~.~~ We first train the ENSO-ASC on the numerous reanalysis ~~dataset~~ samples from 1850.1 to 2015.12 and subsequently on the high-quality remote sensing ~~dataset~~ samples from 1997.12 to 2012.12 for fine-tuning.~~,~~ ~~T~~this procedure is also known as transfer learning. These two datasets are adjusted to the same resolution. The validation period is from 2014.1 to 2020.8 in remote sensing dataset. The gap between fine-tuning set and validation set is used to remove the possible influence of oceanic memory (Ham et al., 2019).

105 It is the first time to design a multivariate air-sea coupler considering energy interactions. W~~w~~e evaluate the ENSO-ASC from three aspects: Firstly, we evaluate the model performance from the perspective of model structure, including the input sequence length, the benefits of transfer learning, multivariate air-sea coupler and the attention ~~components~~weights, and tune the model structure to optimal. Then, we analyze the ENSO forecast skill of the ENSO-ASC from the meteorological aspects, including the contributions of different input physical variables, the effectiveness of forecast lead time, the forecast skill

110 changes with different start calendar months, and the forecast spatial uncertainties. Subsequently, we make the real-world ENSO simulations during the validation period by tracing the evolutions of multiple physical variables. From the experiment results, ENSO-ASC performs better in both *SSIM* and *PSNR* of the forecasted SST patterns, which effectively raises the upper limitation of ENSO forecast~~s~~. The forecasted ENSO events are more consistent with real-world observations and the related Niño indexes have higher correlations with observations than traditional methods and current state-of-the-art deep learning

115 models, which is over 0.78/0.65/0.5 within the lead time of 6/12/18 months for Niño3.4 index. SST and zonal wind are two crucial predictors, which can be considered as the major triggers of ENSO. A temporal heat map analysis illustrates that the ENSO forecasting skills decline faster when making forecasts through June-July-August, and a spatial heat map analysis shows that the forecast errors are more likely to show up over the central tropical Pacific Ocean in longer-term forecasts. At the meanwhile, in the validation period from 2014 to 2020, the multivariate air-sea coupler can capture the latent ENSO dynamical

120 mechanisms and provide multivariate evolution simulations with a high degree of physical consistency: The positive SST anomalies first show up over the eastern equatorial Pacific with the westerly wind anomalies in the western and central tropical Pacific Ocean (vice versa in the La Niña events), which induces Bjerknes positive feedback mechanism. It is worth noting that for the simulation of 2015/2016 super El Niño, ENSO-ASC captures its strong evolutions of SST anomalies over the northeast subtropical Pacific in the peak phase and successfully predicts its very-high-intensity and very-long-duration, while many

125 dynamic or statistical models fail. At the same time, ENSO-ASC can also reduce false alarm rate such as in 2014. From the

4

mathematical expression, the multivariate air-sea coupler captures the spatial-temporal multi-scale oscillations of Walker circulation and ~~handle~~ performs the ocean-atmosphere energy exchange simultaneously, which tries to avoid the interval flux exchange in geoscience fluid programming of traditional numerical climate models. In conclusion, the graph-based multivariate air-sea coupler exhibits not only effectiveness and superiority to predict sophisticated climate phenomena, but 130 also is a promising tool for exploiting the underlying dynamic mechanisms in the future.

The remainder of this paper is organized as follows. Section 2 introduces the proposed multivariate air-sea coupler. Section 3 describes the ENSO deep learning forecast model with the coupler (ENSO-ASC) in detail. Section 4 illustrates the datasets, experiment schemas and result analyses. Finally, Section 5 offers further discussions and summarizes the paper.

2 Multivariate Air-Sea Coupler based on Graph

135 ENSO is the most dominant phenomenon of air-sea coupling over the equatorial Pacific, and many complex dynamical mechanisms modulate the ENSO amplitudes. Bjerknes positive feedback (Bjerknes, 1969) is one of the most significant effects, the processes of which are highly related to the status of Walker circulation. There are energy interactions between the multiple physical variables influenced by Walker circulation every moment, and the ENSO-related SST varieties are greatly affected by such air-sea coupling activities (Gao and Zhang, 2017; Lau et al., 1989; Lau et al., 1996).

140 Many atmospheric and oceanic anomalies can be known as triggers of ENSO events, which establish~~es~~ the Bjerknes positive feedback. The warming SST anomalies propagate to the central and eastern equatorial Pacific gradually. As SST gradually rises, it is virtually impossible for the equatorial Pacific to enter~~s~~ a never-ending warm state. Therefore, some negative feedback will cause turnabouts from warm phases to cold phases (Wang et al., 2017). These negative feedback mechanisms all emphasize air-sea interactions. For example, westerly wind anomalies in the central tropical Pacific Ocean 145 induce the upwelling Rossby and downwelling Kevin oceanic waves, both of which propagate and reflect on the continental boundary, then tend to push the warm pool back to its original position in the western Pacific. From the perspective of ENSO life-cycle, atmospheric/oceanic variables play crucial roles together.

At the meanwhile, during the development and decay phase_s_ of ENSO, there also exist nonlinear interactions between atmospheric and oceanic variables. Wind anomalies are the most obvious and direct response of the ENSO-driven large-scale 150 oceanic varieties, and they will change the ocean-atmosphere heat transmission (Cheng et al., 2019). Once the ocean status changes, the thermal energy contained in the sea will escalate or lose into the air, hindering or promoting the precipitation and surface humidity over the equatorial Pacific. These changes also give feedback on the ENSO.

Meteorological researchers have already identified the key physical processes in ENSO in recent years. If such knowledge can be incorporated into ENSO deep learning forecast modeling subjectively, breaking away from the current limitation of 155 using single predictors, the accuracy of ENSO prediction will promise breakthroughs. In this paper, we choose 6 ~~different~~ ENSO-related indispensable variables from two different multivariate datasets respectively as shown in Table 1, which have strong correlations within the evolution of ENSO events according to Bjerknes positive feedback and other dynamical

5

processes. Furthermore, in order to comprehensively represent the coupl~~inged~~ interactions, a multivariate air-sea coupler $coupler(G)$ is designed to simulate their ~~coupling~~ synergies with an undirected graph $G = (V, A)$ as shown in Fig. 2, where $V = (f_{v_1}, f_{v_2}, ..., f_{v_N})$ represents the vertices of graph and f_{v_i} is the feature of <u>every</u> physical variable~~s~~ v_i ($i = 1, 2, ..., N$). $A \in R^{N \times N}$ is the pre-designed adjacency matrix, where $A_{i,j} = 1$ ($A_{i,j} = 0$) represents ~~there~~ <u>the</u> existing (~~not~~ <u>non-existent</u>~~existing~~) energy interactions between the connected variables v_i and v_j. The variables exchange energ<u>ies</u>~~y~~ simultaneously every moment, and the directions of edges in this graph can be neglected because the energy interactions are two-way (transfer and feedback).
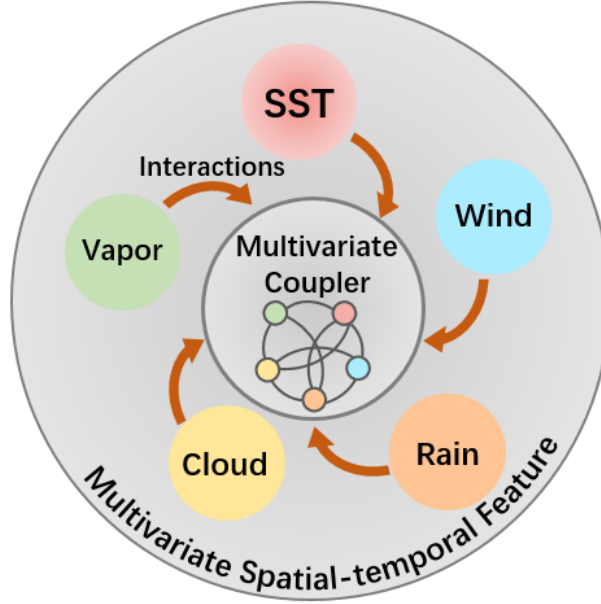

Figure 2: <u>A description of o</u>~~O~~ur proposed multivariate air-sea coupler, which utilizes the spatial-temporal features of multiple physical variables to simulate the energy exchanging simultaneously.

Here, V in G is not the physical variable on a single grid point, but the features of the ~~variable on the~~ entire <u>variable</u> pattern. The reason lies in that: On the one hand, the coupler will pay more attention on the global/local spatial-temporal correlations in the variable fields of ENSO rather than the variations on an isolated grid. On the other hand, the coupler will provide a higher computational efficiency and consume a lower calculation resource for ENSO forecast. The finer and improved couplers, such as designing individual graph for smaller-scale regions and even a single grid, are future considerations.

3 ENSO-ASC: ENSO Deep Learning Forecast Model with the Multivariate Air-Sea Coupler

Inspired by previous ENSO deep learning forecast model, we can define the ENSO forecast as a multivariate spatial-temporal sequence forecast problem as illustrated in Eq. (1),

$$\hat{s} = \mathcal{F}_\theta(s^{scm}), \{sst, uwind, vwind, rain, cloud, vapor\} \subseteq s^{scm}\cancel{s} \qquad (1)$$

where $s^{scm} \in R^{N \times M}$ is N multivariate observations in historical M months ($N = 6$), and $\hat{s} \in R^{N \times H}$ is the prediction result for future H months (H can be also treated as forecast lead time). $scm \in \{Jan, Feb, ..., Dec\}$ (start calendar month) represents the last month in the input series $s^{scm}\cancel{s_{scm}}$. $\mathcal{F}_\theta$ represents the forecast system (θ is the trainable parameters in the system).

In order to ~~combine~~ incorporate the multivariate coupler, we break down the conventional formulization and redefine the multivariate ENSO forecast model as the encoder-coupler-decoder structure shown as Eq. (3) to Eq. (5),

$$\text{encoder: } f_{v_i} = encoder_i(s_i^{scm}) \qquad (2)$$

$$\text{coupler: } f_c = coupler(G) = coupler((f_{v_1}||f_{v_2}|| ... ||f_{v_N}), A) \qquad (3)$$

$$\text{decoder: } \hat{s}_i = decoder_i(f_{v_i}||f_{c_i}) \qquad (4)$$

where the s_i^{scm} ($i = 1,2, ..., N$) represents the individual physical data and f_{v_i} represents the corresponding extracted features by respective encoders. The $coupler(\cdot)$ simulates the latent multivariate interactions on the physical features and the pre-designed interaction graph $A\cancel{G}$, where the operator $||$ represents the concatenation of ~~all~~ features of different physical variables. Then the respective decoders will restore the physical features end-to-end by the coupled multivariate features f_{c_i} and original physical features f_{v_i}, the concatenation of which can be regarded as skip-layer connections. These connections can propagate the low level feature to high levels of the model directly, preserving the raw information and accelerat~~ing~~e feature transfers to some extent. The sub-modules $encoder_i(\cdot)$, $coupler(\cdot)$, and $decoder_i(\cdot)$ form the ENSO-ASC together.

As mentioned before, the strength of the multivariate coupling and the ~~influence~~ effects of multivariate temporal memor~~ies~~y in ENSO ar~~is~~e changing with different input sequence length M and forecast start calendar month scm. In order to ~~weaken~~ grasp such ~~negative impacts~~effects on forecast~~s~~, we design two self-supervised attention weights, $\alpha = (a_1, a_2, ..., a_M)$ and $\beta = (b_1, b_2, ..., b_N)$, in the encoder and coupler respectively to capture the dynamic time series non-stationarity and re-weight the multivariate contributions. The final formulization of forecast model can be written in ~~shown in~~ Eq. (5) to Eq. (7) as shown in Fig. 3, where $\circ$ represents the element-wise multiplication.

$$\text{encoder: } f_{v_i} = \alpha \circ encoder_i(s_i^{scm}) \qquad (2)$$

$$\text{coupler: } f_c = coupler(\beta \circ G) = coupler((b_1 f_{v_1}||b_2 f_{v_2}|| ... ||b_N f_{v_N}), A) \qquad (3)$$

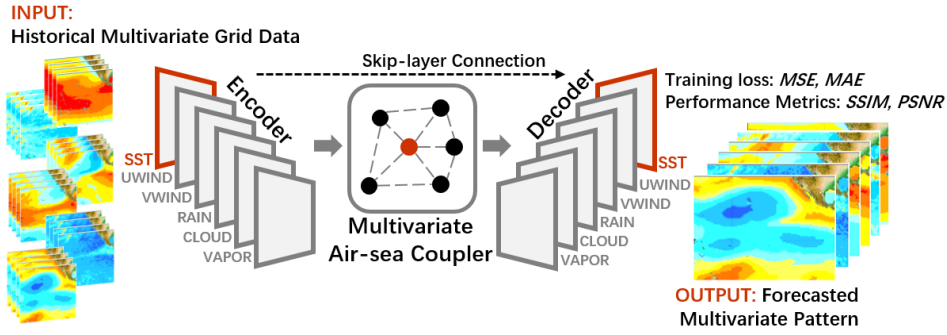$$\text{decoder: } \hat{s}_i = decoder_i(f_{v_i}||f_{c_i}) \qquad (4)$$

Figure 3: The structure of ENSO-ASC. There are 6 encoders for the chosen variables to extract spatial-temporal information and a multivariate air-sea coupler to simulate interactions. After interactions, we design 6 decoders to restore the major variable SST and other variables. The training loss and performance metrics are also displayed.

In addition, there are basically two forecast strategies for ENSO prediction: Ddirect multi-step (DMS) and iterated multi-step (IMS) (Chevillon, 2007). The former means predicting the future H^{th}-month multivariate pattern directly, and the latter means utilizing the forecast output result as the input for future iterated predictions. Figure 4 displays the differences between DMS and IMS. In general, DMS is often unstable and more difficult to train for a deep learning model (Shi and Yeung, 2018). Therefore, we choose IMS to handle chaos data and provide more accuracy predictions, that is, ~~outputting~~ forecasting the next one-month~~'s~~ ($H = 1$) multivariate ~~forecast result~~data in the model, and then using this output as model input to continuously ~~continue forecasting~~ predict the future evolutions. We also design a combined loss function to train our model and use two ~~result~~ spatial metrics to evaluate the forecast results. The intentions and detailed implementations of every part~~s~~ in the model are interpreted as the following sections.
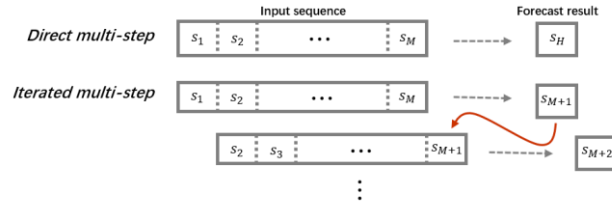

Figure 4: Two common ways in sequence prediction ~~such as~~for ENSO forecast~~s~~: Direct multi-step (DMS) and iterated multi-step (IMS).

3.1 Encoder: Stacked ConvLSTM Layers for Extracting Spatial-Temporal Features

The ENSO evolution has a strong correlation with historical atmospheric/oceanic memory (Zhang et al., 2019b). ENSO deep learning forecast model should be able to simultaneously extract the long-term spatial-temporal features from multivariate geoscience grid data and effectively mine the complicated nonlinearity hidden in the data. Stacked ConvLSTM layers are constructed as the skeleton of the encoder (See orange arrows in Fig. 5) for each chosen physical variable individually.~~Considering the complex spatial-temporal characteristics of ENSO, the encoder is composed of three sub-structures: stacked ConvLSTM layers, 3D max-pool layers and the temporal attention weight.~~
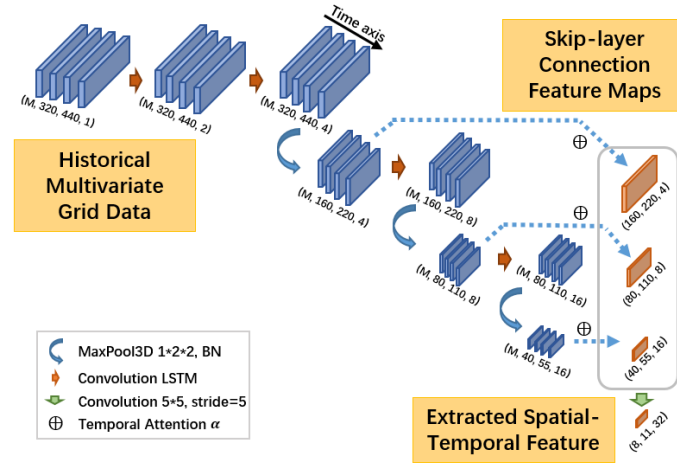
8

Figure 5: A detailed structure for encoder: Stacked ConvLSTM encoder for extracting spatial-temporal features simultaneously. There is also temporal attention ~~component~~ weight for skip-layer connection~~s~~ in the grey box.

~~The ENSO evolution has a strong correlation with historical atmospheric/oceanic memory (Zhang et al., 2019b). ENSO deep learning forecast model should be able to simultaneously extract the long-term spatial-temporal features from multivariate geoscience grid data and effectively mine the complicated nonlinearity hidden in the data. stacked ConvLSTM layers are constructed as the skeleton of the encoder (See Fig. 5) for each chosen physical variable individually.~~

In order to capture the multi-scale spatial ~~remote correlations~~teleconnections ~~at different resolutions~~ in ENSO amplitudes, we ~~take advantage of~~set a 3D max-pool layer between two ConvLSTM layers respectively (as blue arrows in Fig. 5), the stride of which on time-axis is set to 1 to retain the sequence length M. Considering these obtained multi-scale spatial features after every 3D max-pool layers, we design the skip-layer connections shown in the grey box in Fig. 5. These layers ~~pass~~propagate and cascade the raw features of the same variable from its encoder (lower levels) to its decoder (higher levels) directly (See Fig. 9) like dense connection (Huang et al., 2017). Such structure can preserve more details at multi-scale spatial ~~correlations~~teleconnections and also solve the problem of gradient disappearance. In addition, we design the encoders and decoders to be symmetric, ensuring that the feature maps in these connections have the same shape.

Since we set $H = 1$ ~~in~~as the IMS forecast strategy, the feature maps on the encoder all have a time axis, which the decoders do not have. The memory effects on the forecast are mutative with different input sequence length and forecast start calendar month, if we ~~pass~~propagate all time steps' feature maps ~~from~~in the encoder to the coupler and the decoder ~~and the coupled module~~, it is too redundant and even causes over-averaged forecast results, ~~and~~which hinders the descriptions of the special ~~serious~~seasonal ~~interference~~amplitudes ~~will also show up with different start calendar months~~. Therefore, before the skip-layer connections, we first determine the attention weights to dynamically fuse multiple time steps' feature maps in encoder, which can capture the seasonal periodicity hidden in the physical variables and is also called temporal attention weight α (as $\oplus$ symbols in Fig. 5). ~~Meanwhile, this approach can also dynamically extract strongly correlated temporal memories and help improve the prediction skill.~~

9

After obtaining sequential feature maps $T = [t_1, t_2, \ldots, t_m], m = 1, 2, \ldots, M$ from each 3D max-pool layer, we first flatten every time step feature map $t_m \in R^{w \times h \times c}$ by the width w, height h and channel c as $t'_m \in R^{1 \times (w \times h \times c)}$, then cascade ~~connect~~ them together alone time axis as $T' = [t'_1, t'_2, \ldots, t'_m], m = 1, 2, \ldots, M$, where $T' \in R^{M \times (w \times h \times c)}$. We apply Eq. (8) on T' to determine the self-supervised attentive weight $\alpha \in R^M$ for each time step's feature map t_m.

$$\alpha = softmax(W_{\alpha t} \tanh(W_t T' + b_t) + b_{\alpha t}) \tag{8}$$

where $W_t \in R^{d_1 \times M}$ and $W_{\alpha t} \in R^{d_1}$ are transformation matrices, d_1 is a hyper parameter, $b_t \in R^{d_1}$ and $b_{\alpha t} \in R$ are biases. Every dimension in α represents the contribution to the forecast of corresponding time step, ~~so~~ and we use Eq. (9) to fuse the original feature $T = [t_1, t_2, \ldots, t_m], m = 1, 2, \ldots, M$.

$$\tilde{T} = h(\alpha, T) = \sum_{m=1}^{M}(\alpha_m \circ t_m) \tag{9}$$

where $\tilde{T}$ is the aggregated feature map for skip-layer connections, function $h(\cdot)$ represents the summary of element-wise multiplication.

The feature map sizes are described in the Fig. 5 in detail. The sizes of ConvLSTM kernels are all 3×3 and the channel sizes are [2,4,8,16] during forward propagation, where the changes between adjacent layers are smooth and small. The final output (with size of $16 \times 40 \times 55$) of the encoder is generated by a convolution layer of $32 \times 5 \times 5$ with stride 5 and output a feature map with size of $8 \times 11 \times 32$, which ~~can~~ is used to ~~reduce~~ filter the noise derived by such the deep-layer structure.

3.2 Multivariate Air-Sea Coupler: Learning Multivariate Synergies via Graph Convolution

From the perspective of ENSO dynamics, the occurrences of ENSO are accompanied by energy interactions. Based on our formalization and chosen physical variables, we define the corresponding adjacency matrix $A \in R^{6 \times 6}$ and degree matrix $D \in R^{6 \times 6}$ (the diagonal matrix, the value on the diagonal indicates the number of other vertices connected to this vertices) as Eq. (10), which means all physical variables have ~~coupling~~ coupled interactions with each other (full-connected).

$$A = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}, D = \begin{bmatrix} 6 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 6 \end{bmatrix} \tag{10}$$

In practical implementation mathematically, we use graph Laplacian matrix L to normalize the energy flow of original adjacency matrix A as Eq. (11), where I_n is an identity matrix with the order $n \times n$. L can be considered as the directions in which the excess unstable energy will propagate to other variables when the entire system is perturbed (such as external wind forcing).
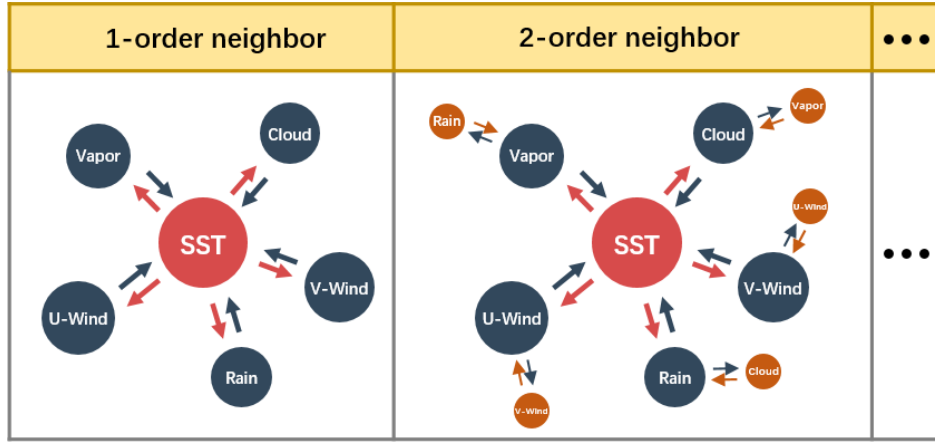
$$L = I_n - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \tag{11}$$

Figure 6: Multivariate coupl~~ing~~ed interactions within K-order neighbours (~~T~~taki~~ng~~e SST as the center).

Meanwhile, the interactions between ENSO-related variables are cascaded, which means that the effects between physical variables are multi-order as depicted in Fig. 6. For example, the precipitation anomalies affect the wind anomalies, which in turn affect the evolutions of~~s~~ SST, as depicted in Fig. 6 (right). According to the properties of Laplacian matrix, L^k is employed to determine the cascaded interactions between k-order neighbors. So, if we consider K-order ~~neighbors~~effects, the whole process is defined by Eq. (12),

$$P^{(l+1)} = \sigma(\sum_{k=1}^{K} \Theta^{(k)} L^k P^{(l)}) \tag{12}$$

where $\Theta^{(k)}$ represents the latent trainable multivariate interactions. K represents the truncated order of ~~neighbours~~ effects concerned. $P^{(l)}$ represents the input features before coupling interactions and $P^{(l+1)}$ represents the coupled features. Each row in both $P^{(l)}$ and $P^{(l+1)}$ represents the same variables. Activation function σ increases the nonlinearity. Figure 7 illustrates the above process mathematically, which is named as the K-order graph convolution layer.
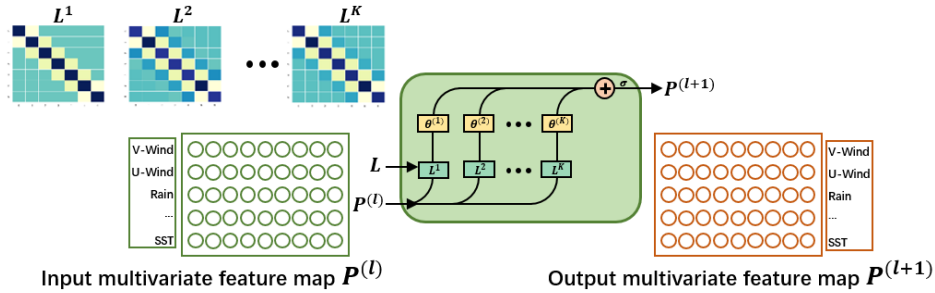

Figure 7: K-order g~~G~~raph convolution layer. K is the truncated order and L^k represents the interactions between k-order neighbours. ~~and e~~Each row in $p^{(l)}$ and $p^{(l+1)}$ represents features of different physical variables, and their positions are not changed during forward propagation.

~~In deep learning,~~The K-order graph convolution network (GCN~~, Bruna et al., 2013~~) ~~is applied to implement~~in Eq. (12) is actually the higher-order extension of original GCN (Bruna et al., 2013). Furthermore, we use Chebyshev polynomial $T_k(\tilde{L})$

11

to approximate the higher-order polynomial $[L^1, L^2, ..., L^k]$, $k = 1,2, ..., K$ for accelerate calculation, where $\tilde{L} = 2L/\lambda_{max} - I_n$ scales L within $[-1,1]$ for satisfying Chebyshev polynomial and λ_{max} is the maximum Eigen value of L (Hammond et al., 2011; Defferrard et al., 2016). This approximation ~~makes~~ accelerates the~~ ~~calculation of the K-order ~~graph convolution~~GCN ~~more promising~~ by reducing the computational complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(K|\varepsilon|)$ ($|\varepsilon|$ is the edge count in the graph). Based on such neural structure, we construct the multivariate air-sea coupler (ASC) to learn synergies related to ENSO as Fig. 8.
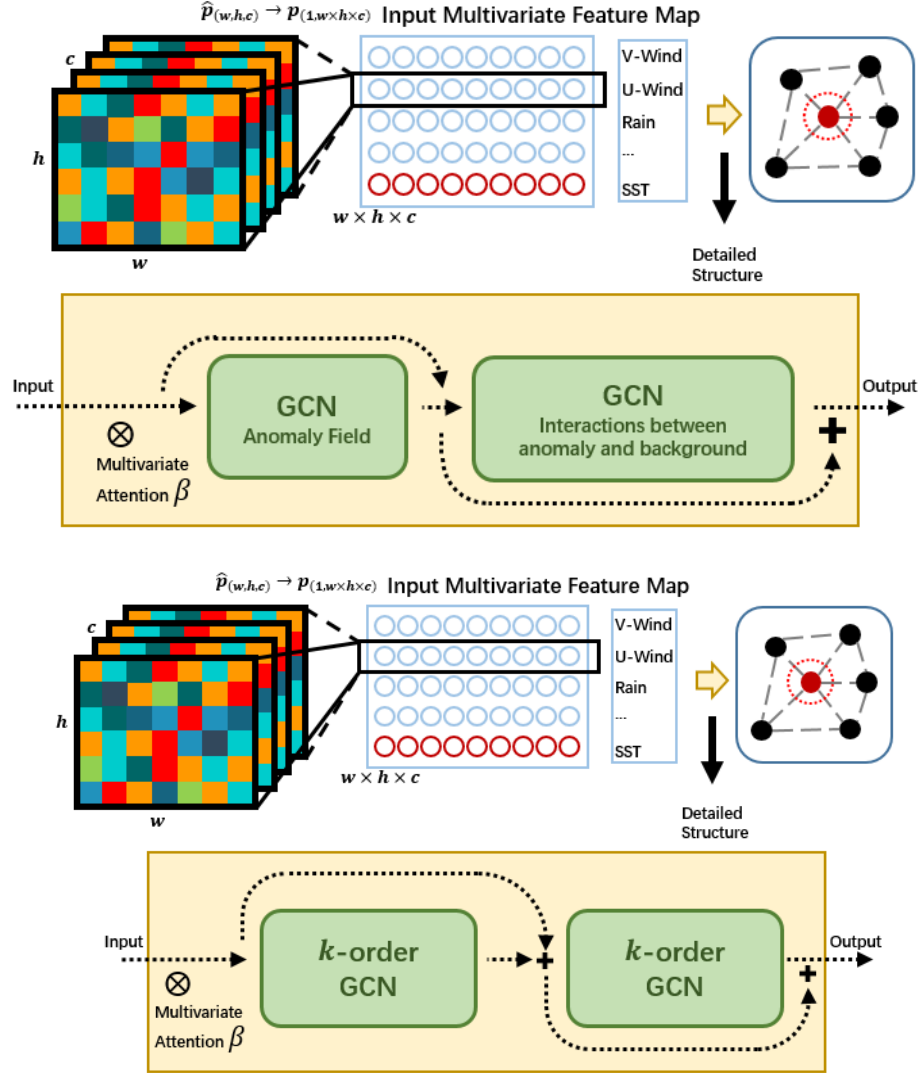

Figure 8: A detailed structure for multivariate air-sea coupler between encoder and decoder: Pre-processes for input (upper row), and dual-layer structure for ~~simulating interactions between anomalous and background fields~~residual learning (lower row). There is also a multivariate attention ~~component~~ weight in the coupler.

After obtaining the spatial-temporal features (Such as the colored feature maps in Fig. 8) from multivariate encoders respectively, we first flatten and cascade them as $P^{(l)}$ like the blue box in Fig. 8~~(See Fig. 8 (upper row))~~. As mentioned above,

12

305   each row of $P^{(l)}$ represents different variables. $P^{(l)}$ is marked as multivariate feature map and acts as the input of the coupler. The coupler is designed as a dual-layer summation structure like the yellow box in Fig. 8(See Fig. 8 (lower row)). The input for the second layer is the sum of the input and the output of the first layer, and the output of the second layer is determined by the weighted fusion of the outputs of these two layers as the manner designed by Chen et al. (2019), ,which is the residual learning to enhance the generalization ability of the network (He et al., 2016). The output of the multivariate coupler is

310 determined by the weighted fusion of the outputs of these two layers as Chen et al. (2019).

     Because the variables contribute differently to the ENSO forecast especially in different start calendar months, we propose the multivariate self-supervised attention weight for determining the weights effects for the input physical variables (as $\otimes$ symbols in Fig. 8). Before $P^{(l)}$ passing into the multivariate coupler, the weight $\beta \in R^N$ for each variable is determined by Eq. (13).

315   $\beta = softmax(W_{\beta p}\ tanh(W_p P^{(l)} + b_p) + b_{\beta p})$ (13)

where $W_p \in R^{d_2 \times N}$ and $W_{\beta p} \in R^{d_2}$ are transformation matrices, d_2 is a hyper parameter, $b_p \in R^{d_2}$ and $b_{\beta p} \in R$ are biases. Then we use Eq. (14) to calculate the modulated multi-physical variables feature map $P^{(l)}$, where $g(\cdot)$ represents the element-wise multiplication.

  $P^{(l)} = g(\beta, P^{(l)}) = \beta \odot P^{(l)}$ (14)

320   In the multivariate air-sea coupler, the corresponding locations of physical variables on the input feature map and output feature map are fixed. For example, if we set the SST feature in the last row of the input multivariate feature map of the coupler, the SST feature will be in the last row of the output multivariate feature map (Seeas shown in Fig. 7), which and will be passed propagated to the decoder for pattern prediction later.

3.3 Decoder: End-to-end Learning to Restore the Forecasted Multivariate Patterns

325 ENSO evolution is considered as a hydrodynamic process. Meteorologists usually use linear methods (such as EOF or SVD method) to extract features, then analyze the potential characteristics and predict the future evolution of ENSO. In these methods, complex dynamical processes are usually simplified to facilitate calculations while unknown detailed processes are not comprehensively revealed or even neglected, which leads to low prediction accuracy. Therefore, we use the end-to-end learning to restore the evolutions of multi-physical variable patternss. The multi-scale spatial-temporal correlations should be

330 also considered in this process, so the decoder consists of stacked transform-convolution layers and up-sampling layers.
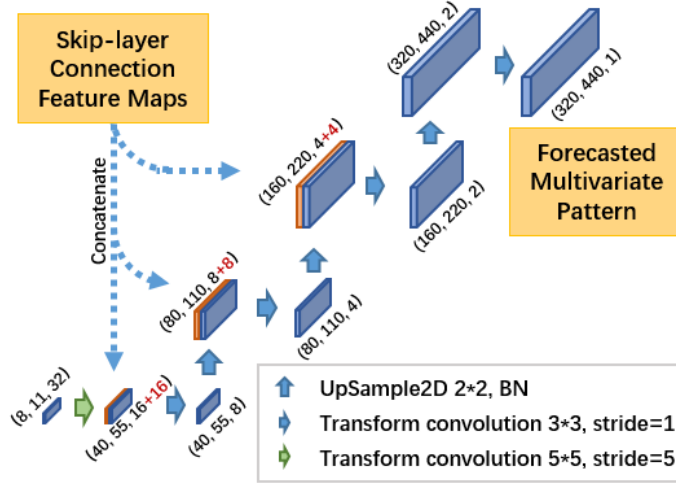
Figure 9: A detailed structure for decoder, skip-layer connections ~~(the red channel cascaded to the feature maps)~~ from encoder for helping end-to-end learning to restore the forecasted patterns at different spatial scales.

From the output feature map of the multivariate air-sea coupler, we pick up the corresponding row (~~T~~taking SST as an example) as P_{SST} (~~S~~such as red row and red circle in Fig. 8), and reshape it into original shape $P'_{SST} \in R^{w \times h \times c}$. Then P'_{SST} is gradually amplified and restored in the decoder ~~with~~ by the stacked transform-convolution layers and up-sampling layers (See Fig. 9). Skip-layer feature maps from encoder are cascaded with corresponding layers with the same shape. The sizes of convolution kernels are all 3×3 which is the same with that in encoder, and the channel sizes are [16,8,4,2,1] to shrink the channel size gradually during forward propagation.

3.4 Loss Functions for model training

Our goal is to predict the evolutions of multiple physical variables (~~marked~~ Marked as $\hat{s}$) as accuracy as possible compared with the real-world observation s. Therefore, we combine two different measurements together as the loss function ℓ in Eq. (15) to ensure the result precision~~spatial consistency~~ of ~~restored~~ multivariate patterns grid by grid,

$$\begin{cases} MSE = \frac{1}{N\Omega} \sum_N \sum_\Omega (\hat{s}_{i,j} - s_{i,j})^2 \\ MAE = \frac{1}{N\Omega} \sum_N \sum_\Omega |\hat{s}_{i,j} - s_{i,j}| \\ \qquad \ell = MSE + MAE \end{cases}, (i,j) \in \Omega \tag{15}$$

where N is the number of variables and Ω represents ~~all~~ the number of grid points ~~of~~ for every physical pattern, (i,j) represent different latitude and longitude. ℓ is consist of the sum of MSE and MAE, where MSE is used to preserve the smoothness of the forecasted patterns and MAE is used to retain the peak distribution of all grid points.

3.5 Metrics to evaluate the forecast results

According to the loss function, the calculation processes in Eq. (15) ~~are~~ mainly focus~~ed~~ on the comparisons of single grid in

350 fields. However, the detailed spatial distributions of every physical variable, such as the ~~location of the~~ max value region ~~for~~of SST and ~~the anomaly regions of~~ wind anomalies, are more important in the ENSO forecast. Therefore, we especially use ~~these~~ the following two common ~~quality~~ spatial metrics for the forecasted ~~SST~~ patterns to evaluate the ENSO forecast skill: $PSNR$, and $SSIM$ as Eq. (16) and Eq. (17),

$$PSNR = 10 \cdot log_{10}(\frac{MAX^2}{MSE}) \tag{16}$$

$$\begin{cases} luminance = \frac{2\mu_{\hat{s}}^2\mu_s^2+c_1}{\mu_{\hat{s}}^2+\mu_s^2+c_1} \\ contrast = \frac{2\sigma_{\hat{s}s}+c_2}{\sigma_{\hat{s}}^2+\sigma_s^2+c_2} \\ structure = \frac{\sigma_{\hat{s}s}+c_3}{\sigma_{\hat{s}}\sigma_s+c_3} \\ SSIM = luminance^a \cdot contrast^b \cdot structure^c \end{cases} \tag{17}$$

In $PSNR$, MAX is the maximum ~~of~~ among all grids. In ENSO-ASC, before the historical multivariate data ~~input~~ is propagated into the model, we first normalize them in the range [0,1] as Eq. (18). Therefore, MAX is set to 1.

$$x^* = \frac{x-x_{min}}{x_{max}-x_{min}} \tag{18}$$

~~In~~ $SSIM$ is a combined metric of luminance, contrast and structure between two patterns~~.~~ $\mu_{\hat{s}}$ (μ_s) is the average for $\hat{s}$ (s), and $\sigma_{\hat{s}}$ (σ_s) is the corresponding standard deviations. $\sigma_{\hat{s}s}$ represents the covariance, $a = b = c = 1$ for fair measurement of every ingredient of $SSIM$, c_1, c_2, and c_3 are all ~~with~~ trivial values for preventing the denominator from being 0.

Besides these two metrics, the correlations between the calculated and the official Niño~~3.4~~ index~~es~~ will ~~also~~ be also used to evaluate forecast skills.

4 Experiments Results and Analysis

4.1 Dataset Description

After the deep learning model structure is determined, the ~~quality and~~ quantity and quality of training dataset affect the forecast performance decisively. As the improvement of observation ability, there are growing ways to provide multiple real-world observations, such as remote sensing satellite, buoy observation, which is more and more beneficial to build our ENSO forecast model. However, one of the biggest limitations in high-quality climate dataset is that the real-world observation period is too short to provide adequate samples. For example, extensive satellite observations ~~may~~ have started in 1980s, and the number of El Niño that occurred ever since then is also small, which is easy to cause the under-fitting of the deep learning network.

Table 1: Multi-physical variables in the corresponding two datasets

NOAA/CIRES		REMSS	
Variable	Description	Variable	Description
SST	Sea surface temperature	*SST*	Sea surface temperature

PWAT	Precipitation Water (Atmospheric Column)	RAIN	Rate of liquid water precipitation
CWAT	Cloud Water (Atmospheric Column)	CLOUD	Total cloud liquid water (Atmospheric Column)
RH	Surface relative humidity	VAPOR	Total gaseous water (Atmospheric Column)
UWIND	Surface zonal wind speed	UWIND	Zonal wind speed
VWIND	Surface meridional wind speed	VWIND	Meridional wind speed

Note: The reanalysis dataset is provided by NOAA/CIRES, which is from 1850.1 to 2015.12 with 2 by 2 degree, and the remote sensing dataset is provided by REMSS, which is from 1997.12 to 2020.8 with 0.25 by 0.25 degree. In REMSS, UWIND and VWIND is collected from REMSS/CCMP (https://rda.ucar.edu/datasets/ds131.2/index.html), and other variables are collected from REMSS/TMI (1997-2012, http://www.remss.com/missions/tmi/) and REMSS/AMSR2 (2012-2020, http://www.remss.com/missions/amsr/). We try to choose physical variables in NOAA/CIRES with the same meaning with that in REMSS, such as CWAT and CLOUD, RH and VAPOR. Limited by these two datasets, some variables can only find the closest match though they describe the different characteristics in ocean-atmosphere cycle, such as PWAT and RAIN.

To greatly increase the quantity of training data, we utilize the transfer learning technique to train our model with long-term climate reanalysis data and high-resolution remote sensing data togetherprogressively. These two datasets both provide multivariate global global gridded dataobservations. The reanalysis data is supported by NOAA/CIRES (https://rda.ucar.edu/datasets/ds131.2/index.html), which is a 6-hourly multivariate global climate dataset from 1850.1 to 2015.12 with 2° resolution. The remote sensing data is obtained from Remote Sensing System (REMSS, http://www.remss.com/), which is a daily multivariate global climate dataset from 1997.12 to 2020.8, and the resolution is much higher than reanalysis data with 0.25°. According to our chosen physical variables, we obtain the corresponding sub-datasets, and all the variables are preprocessed monthly-averaged. The detailed dataset descriptions are shown in Table 1. Note that we try to choose physical variables in NOAA/CIRES with the same meaning with that in REMSS, such as CWAT and CLOUD, RH and VAPOR. Note that sSome variables can only find the closest match in these two datasets though they describe the slightly different characteristics in ocean-atmosphere cycle, such as PWAT and RAIN.

In addition, we also collect the historical NiñoNiño3, NiñoNiño4 and NiñoNiño3.4 indexes data from China Meteorological Administration National Climate Centre (https://cmdp.ncc-cma.net/). We pick up the records from 2014.1 to 2020.8 for the result analysis of following experiments.

The major active region of ENSO is concentrated in the tropical Pacific, so we crop the multivariate data with the region (N40°N-S40°S, E160°E-W90°W) as the geographic boundaries of ENSO-ASC, which covers NiñoNiño3 and NiñoNiño4 regions. The reanalysis data has the size (40, 55) for every single-month variable, and the remote sensing data has the size (320, 440). In order to unify and improve dataset quality, we use Bicubic interpolation (Keys, 1981) to enlarge the reanalysis data by 8× magnification and soft-impute algorithm (Mazumder et al., 2010) to fill up missing values in both datasets. We

train the model first on the whole reanalysis dataset and subsequently on the remote sensing dataset from 1997.12 to 2012.12 for fine-tuning. The samples from 2014.1 to 2020.8 in remote sensing dataset is are considered as validation set. There is a one-year gap between fine-tuning set and validation set to reduce the possible influence of oceanic memory.

4.2 Experiment Setting

We train and evaluate the ENSO-ASC on a high-performance server. Based on our proposed model, some hyper--parameter settings are determined based onreferred to the existing calculation computing resources as following: $K = 4$, $d_1 = d_2 = 16$, which is the optimal parameter combination after extensive experiments (This process has been ignored because it is not the focus in this paper). Adjacency matrix A and corresponding Laplacian matrix $\tilde{L}$ are designed as Section 3. All the following analyses are based on the stable results through extensive repeated experiments.

We evaluate the ENSO-ASC from three aspects. Firstly, according to our proposed ENSO forecast formalization in Eq. (5) to Eq. (7), there are several factors that may influence the performance from the perspective of model structure: Tthe input sequence length M, the multivariate coupler $coupler(\cdot)$, the attention weights α and β, and the benefits of transfer training. We design some comparison experiments to investigate the model performance and determine the optimal model structure. A comparison with the other state-of-the-art models is also included. Secondly, we evaluate the ENSO forecast skill of the ENSO-ASC from the meteorological aspects in according to Eq. (5) to Eq. (7): Tthe contributions of different input physical variables V in the pre-designed coupling graph G, the effectiveness of forecast lead time month in IMS strategy, the forecast skill with different start calendar month scm, and the spatial uncertainties in a longer-term forecasts. Finally, we forecast the real-world ENSO over the validation period and compare our results with the observations.

4.3 Evaluation of Model Performance

4.3.1 Influence of the Input Sequence Length

Input sSequence length M is very important for forecasting model, due to the rich spatial-temporal contained in it. In general, the longer sequence length M, the better the ENSO forecast skill. However, it longer input sequence will also increase the computational burden of the deep learning model and strengthen raise the quality requirements for the collected geoscience data quantity/quality in the training and calculation of complex deep learning network, especially under such a high resolution of our model. Therefore, the balance between forecast performance and efficiency must be considered. Our proposed model uses IMS to forecast the ENSO evolution in the next future month, so wWe gradually increase the sequence length M to detect the changes in forecast skills. Figure 10 displays the results. As the sequence length gradually increases, two metrics grow become better (larger). When the sequence length is greater than 3 months, the growth rate slows down., Wwhile the sequence length is less than 3 months, the forecast skill increases rapidly.
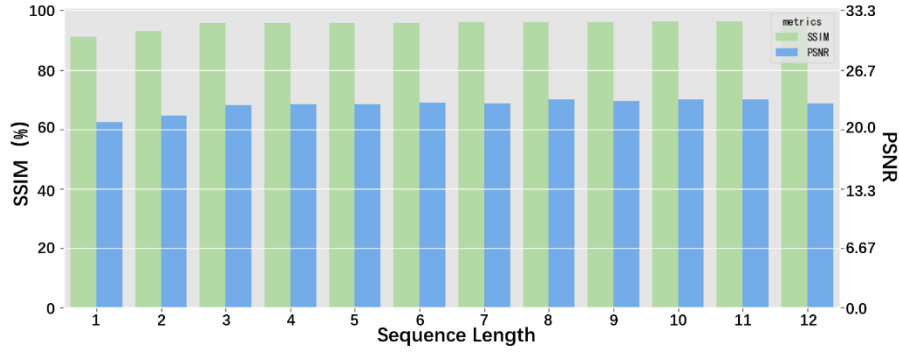
17

Figure 10: The performances of the ENSO-ASC when the input sequence length increases under IMS forecast strategy.

It is obvious that the increase of sequence length cannot lead to an unlimited improvement in forecast skill. In ENSO-ASC, making predictions with the previous 3 months' multivariate data is a more efficient choice. In fact, lots of successful works imply that climate deep learning model does not require a longer input sequence to make skilful predictions, such as using previous 2 continuous time-step data to estimate the intensity of tropical cyclone (Zhang et al., 2019a) and, using previous 3-month SST ocean heat content and wind data to predict ENSO evolution (Ham et al., 2019). A long-term temporal sequence must contains strong trends and periodicities, but the underlying chaos is more dominant, which seriously hinders the prediction. The subsequent experiments will all apply the historical 3-month multivariate sequence ($M = 3$) as model input data.

4.3.2 Benefits of the Transfer Learning

For the all model training, we use the transfer learning approach to overcome the insufficient sample challenge and obtain the optimal trained model. More specifically, we first train the model on reanalysis dataset with 90 epochs and subsequently on remote sensing dataset until total convergence (about 110 epochs). Here, 90 epochs are enough for training the ENSO-ASC on the reanalysis dataset until convergence, because the interpolated reanalysis data is smoother and lacks of details, which leads to an easy-training. In order to verify the benefit of the transfer learning, wWe also make comparative experiments by only training our model on remote sensing dataset. We set the input sequence length at 3 and train the model repeated times to obtain a stable result. The training process needs more epochs (such as 200 epochs), because tThe original remote sensing dataset contains much more detailed high-level climate information, so the model need more epochs (such as 200 epochs) to train.
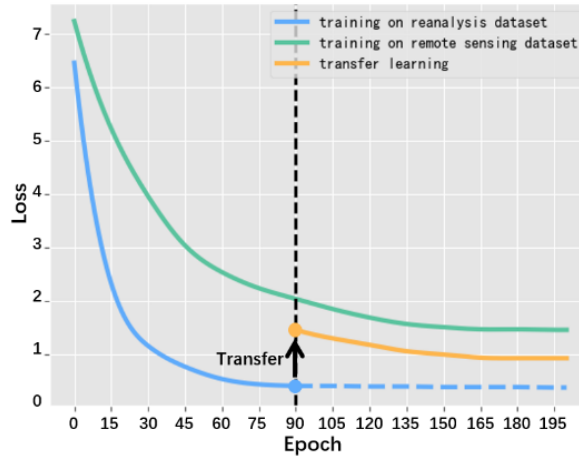
18

Figure 11: The loss changes when training with only reanalysis dataset (blue line), with only remote sensing dataset (green line) and transfer learning on these two datasets in order (black arrow and yellow line after 90 epochs).

450   ~~We also make comparative experiments by only training our model on remote sensing dataset. We set the input sequence length at 3 and train the model repeated times to obtain a stable result.~~ The averaged loss changes for different training sets are~~is~~ depicted in Fig. 11. We can see that when training with the reanalysis dataset, the loss drops quickly, while training with the remote sensing dataset, the model converges slowly and the loss is large. After using transfer learning ~~approach~~, the loss on remote sensing dataset are improved at least 15%, which demonstrates that the systematic errors of ENSO-ASC are indeed corrected to some extent.

455   Comparing with remote sensing dataset, training with reanalysis dataset always yields a much smaller loss. It is due to the smoothness and lack-of-details of the reanalysis dataset as above mentioned, which leads to the model can learn the characteristics more easily. However, the high-resolution remote sensing dataset reflects the real-world ~~conditions~~ status more accurately, which contains more comprehensive and nonlinear details and amplitudes under a high resolution. If we have ~~enough~~ efficient remote sensing data, the forecast skill will be further improved.

460   ### 4.3.3 Effectiveness of the Multivariate Air-sea Coupler

We subjectively incorporate priori ENSO coupled interactions into the graph-based multivariate coupler and select 6 critical physical variables as the predictors of the ENSO-ASC. The formalization not only treats each physical variable as a separate individual, but emphasizes the nonlinear ~~connections~~ interactions between them. However, it is not clear whether such graph formalization is the reason for the improvement of ENSO forecast performance. In order to validate the effectiveness of our
465   proposed formalization, we design two other deep learning couplers for ENSO forecast with the same datasets and transfer learning, and then compare the performance with the ~~original model~~ENSO-ASC.
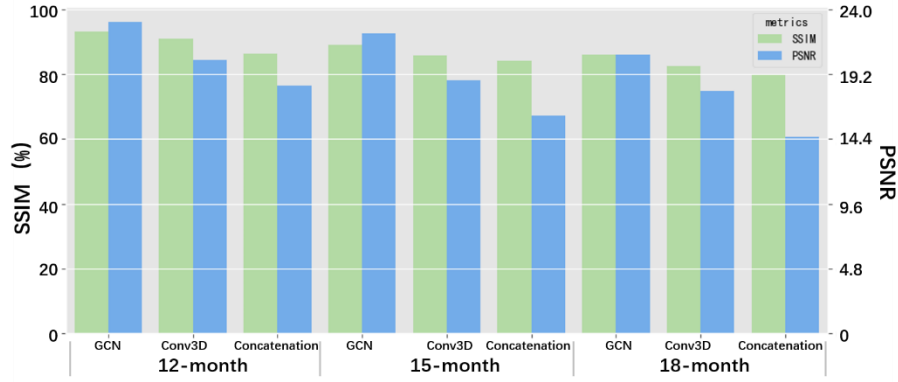
19

Figure 12: The performances of the ENSO-ASC when replacing the multivariate air-sea coupler with other deep learning structures.

The first coupler is to replace GCN with a dual-layer 3D-convolution block, which treats all variables as a whole system and ignores the specific directions and neighbour-orders in coupling~~ed~~ interactions between them. The second coupler just replaces the GCN with the concatenation of features from multivariate encoders, which treats the multiple variables as the~~a~~ channel stacking (or data overlay) and simply extracts global features of them together (the cascaded ~~concatenated~~ multivariate features are ~~passed~~ propagated ~~into~~ to the ~~multivariate~~ decoders for every variable directly). The results are illustrated in Fig. 12. Obviously, our graph formalization achieves the best performance with measurements of *SSIM* and *PSNR*, the Conv3D coupler is slightly worse. The results indicate that using a graph to simulate multivariate interactions is a more reasonable approach, which can learn more ENSO-related dynamical interactions and underlying physical processes than other formalizations. Besides this, the comparative experiments also exhibit some inspiration: When building a climate forecast deep learning network that incorporates physical mechanisms, it is necessary to ~~select~~ customize a suitable structure to ~~support~~ represent and ~~express~~ reflect specific mechanisms mathematically.

We design a full-connected adjacency matrix A in GCN coupler, which means we consider that all physical variables have interactions with each other. Conv3D also entirely extracts the features of all variables. Under this circumstance, why GCN coupler has better performance than Conv3D coupler? From the perspective of mathematics, GCN will consider the pair-wise coupling between variables and learn the features of every coupled interaction according to the hidden nonlinearity in samples individually as Eq. (12). But the Conv3D coupler inherits the characteristics of the global-shar~~inge~~ and local-connect~~ion~~ ~~parametric designs~~ in classic convolution, which rather treats all variables equally and lack descriptions of the special interactions~~treats different variables as different channels in images, lacking more detailed coupled information~~.

4.3.4 Effects of Attention Weights

We customize two attention ~~components~~ weights in the model to dynamically ~~overcome~~ represent the ~~problem~~ effects~~of negative effects~~ of different temporal memories and multiple variables. Here, we analyze the influences of two proposed self-supervised attention weights by removing one of them from ENSO-ASC, the results are illustrated in Fig. 13. The results

20

suggest the prediction skill will decline when one of the attention ~~techniques~~ weights is removed. More specifically, the reduction of performance is larger when the multivariate attention is removed for shorter forecasts (less than about 9 months), ~~or~~ and when the temporal attention is removed for longer forecasts (more than about 15 months). It is because of higher multivariate correlations and lower temporal non-stationarity in the short term. But the temporal memory effects dominant ~~in the long term~~the long-term evolution.
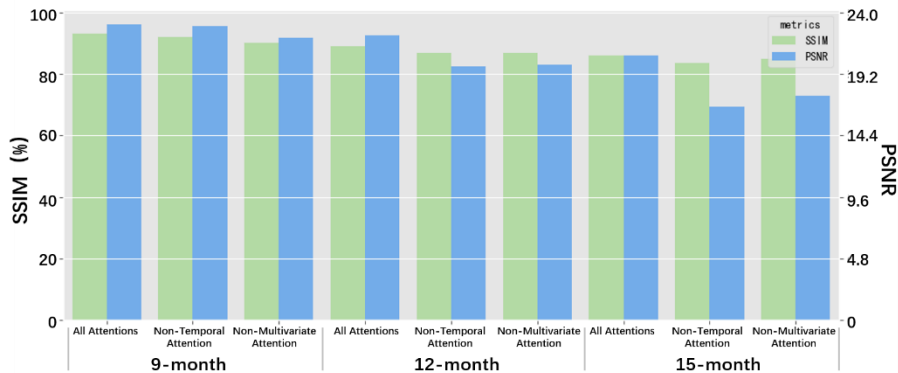

Figure 13: The performance~~s~~ of the ENSO-ASC when removing one of the attention ~~components~~weights.

In fact, due to the self-supervised attentive weights α and β, ~~even~~ though ~~when~~ the multivariate graph and model structure are fixed, the ~~prediction~~ forecast skills will not change too much with the difference start calendar months and variable combination (See Section ~~5.2.3~~4.4). ~~Because of the IMS forecast strategy, the prediction with the different start calendar months will share one common model.~~However, ~~i~~If these two ~~components~~ weights are unset, the model will not be able to distinguish the contributions of multivariate oceanic memories in different ~~situations~~forecast start month adaptively, seriously misleading the forecast~~s~~.

Indeed, it may be a better choice for ENSO forecast to establish and filter the optimal model ~~individually~~ for different start calendar months, forecast lead time and ~~with~~ various ~~variables~~predictors, but it also means higher resource and time consuming. These two attentive weights α and β can reasonably prune the model within the acceptable range of prediction errors, ~~which is more like the traditional dynamic forecasting manner.~~. In operational forecasts, separate modelling ~~methods for different scenarios~~ can be used to pursue higher accuracy and skills.

4.3.5 Comparison with Other State-of-the-art ENSO Deep Learning Models

We compare the ENSO-ASC with other state-of-the-art data-driven ENSO forecast models, including: (1) Convolutional Neural Network (Encoder-decoder structure with 12 layers, which has the same trainable layer number with our model); (2) Long Short-Term Memory Network (6 LSTM layers and final fully connect layer); (3) ConvLSTM Network (CL-6 means a 6-layer structure, CL-12 means a 12-layer structure); (4) 3D-Convolution coupler to simulate multivariate interactions as mentioned above (Conv3D). In order to ensure the fairness of the comparison, we utilize the same input physical variables, training/validation datasets and training criteria for the above models, then train them via transfer learning with plenty of

21

epochs to achieve their optimal performances. Table 2 displays the comparative results with 12-/15-/18-month forecast. In general, the forecast models considering ENSO spatial-temporal correlations (e.g. ConvLSTM, Conv3D) outperform the basic deep learning models (e.g. CNN, LSTM), which implies that the complicated network structures can mine the sophisticated rules dependencies deeply hidden in long-term ENSO forecast evolutions more effectively. As the lead time increases, the

520 performances of models gradually decreases. However, the ENSO-ASC still maintains high accuracy and is always better than other models with an increment improvement of about 5%, which indicates the superiority of our model.

Table 2: Performance comparisons with other state-of-the-art deep learning models

Model	12-month SSIM / PSNR	15-month SSIM / PSNR	18-month SSIM / PSNR
CNN	86.32 / 17.47	83.97 / 14.40	79.85 / 12.20
LSTM	88.57 / 18.08	84.19 / 15.41	81.59 / 13.58
CL-6	88.70 / 18.37	84.36 / 16.25	81.73 / 14.04
CL-12	89.78 / 19.45	84.74 / 17.34	82.03 / 15.37
Conv3D	90.93 / 20.16	85.59 / 18.01	82.50 / 15.98
ENSO-ASC	92.65 / 22.05	90.31 / 20.97	87.53 / 17.17

Considering the calculation ingredients of *PSNR* and *SSIM* according to Eq. (11): The calculation of *PSNR* only contains *MSE*, which is the metric for individual grids, while *SSIM* measures the spatial characteristics and distributions of two patterns

525 from many correlation coefficients, which can represent a measurement for evolution tendency and physical consistency to some extent. Based on the above analysis, Table 2 also indicates that the forecast skill of the forecast results of ENSO-ASC exhibit s an excellent physical consistency big improvement overbeyond other models, the *SSIM* of which is much better, especially in the longer lead time. In addition, the ENSO-ASC pays more attention on the detailed spatial distributions, which is beneficial for the further analysis of ENSO dynamical mechanisms.

530 On the other hand, the ENSO-ASC is the first attempt to forecast ENSO at such a high resolution (0.25 °). Despite the difficulty of training increases, our model still achieves good results. Interestingly, though the ENSO-ASC involves the most trainable parameters, its convergence epoch is one-fifth on average of other forecast models.

4.4 Analysis of ENSO forecast skill

4.4.1 Contributions of Different Predictors to the Forecast Skill

535 The superiority of our proposed model derives from the graph formalization, and the special multivariate coupler can effectively express the processes of synergies between multi-physical variables. From another perspective, the improvement of the forecast skill is not only benefited from graph formalization, but also due to the utilization of multiple variables highly related to ENSO compared to using limited variable to predict ENSO as previous works. For ENSO forecast, SST is definitely the most critical predictor. Besides SST, other variables have different contributions to the forecast results. Therefore, we

540 design an ablation experiment by removing one aone of predictors each time from our proposed model and detect the reduction

of forecast skill (Table 3 above). At the meanwhile, we also add one extra predictor (from surface air temperature, surface pressure and ocean heat content respectively) into our proposed model to investigate the improvement of forecast skill (Table 3 below). Here, the input sequence length is still set to 3 ~~Meanwhile, we also add a predictor each time into our proposed model to detect the improvement of forecast skill (Table 3 below). Here, the input sequence length is still set to 3~~.

Table 3: Model performance when one existing variable removed or one extra variable added

Removed variable	12-month SSIM / PSNR	15-month SSIM / PSNR	18-month SSIM / PSNR
-	92.65 / 22.05	90.31 / 20.97	87.53 / 18.17
RAIN	91.46 / 21.34	88.74 / 18.32	85.86 / 17.35
CLOUD	91.53 / 21.65	88.81 / 18.54	85.93 / 16.16
VAPOR	91.52 / 21.65	88.82 / 18.53	85.92 / 16.16
UWIND	90.08 / 20.93	87.03 / 17.81	83.72 / 13.58
VWIND	91.47 / 21.62	88.65 / 18.42	85.31 / 15.07
Added variable	12-month SSIM / PSNR	15-month SSIM / PSNR	18-month SSIM / PSNR
Surface Pressure	92.74 / 22.13	90.33 / 20.99	87.64 / 17.26
Surface Air Temperature	92.75 / 22.15	90.40 / 21.07	87.71 / 17.25
upper ocean heat content	92.98 / 22.14	90.45 / 21.10	87.79 / 17.34

Table 3 (above) shows that when a variable is removed from the input of the deep learning model, the ENSO forecast skill will be reduced. More specifically, when the zonal weed speed (UWIND) is removed, the reduction is the largest. From the perspective of ENSO physical mechanism, zonal wind anomalies (ZWA) always play a necessary role and are even considered as the co-trigger or driver of ENSO events. As an atmospheric variable, ZWA often gives a direct feedback on oceanic varieties with a shorter response time than oceanic memory. ENSO-ASC uses historical 3-month multivariate data to predict ENSO evolution, which is a quite short sequence length. Under such sequence length, wind speed (including u-wind and v-wind) has a relatively high correlation with SST. In addition, RAIN is another variable that slightly affects the forecast. This is because the precipitation process has a straightforward contact with the sea surface, and the energy transfer is easier.

Table 3 (below) ~~shows~~indicates that the model performance improves a little when ~~when~~ adding surface air temperature/surface pressure/ocean heat content ~~each time~~ into the multivariate coupler~~, the results exhibit a little improvement on the performance~~. This is because that the multivariate graph with existing variables in the ENSO-ASC can almost describe a relatively complete energy loop in Walker circulation, so the effects of the extra added variables to the ~~forecast of~~ ENSO forecasts are not obvious. It is worth noting that ~~Here one particular point to be pointed out is~~ the input sequence length should be longer ~~that~~ when ~~adding~~feeding the ocean heat content into the multivariate coupler, ~~the input sequence length should be longer~~ because this predictor is with long memory (Ham et al., 2019; McPhaden, 2003, Jin, 1997, Meinen and McPhaden,

23

2000). However, as when the input sequence length is set to 9 varies from 3 to 9 months, the forecast skills of ENSO-ASC have not changed much of ENSO-ASC is actually not much different. This is mainly caused because that the global spatial teleconnections and temporal lagged correlations by Walker Circulation and ocean waves (such as Kelvin and Rossby Waves) (Exarchou et al., 2021 and Dommenget et al., 2006) are not caught in the model, the input region of which mainly covers the equatorial Pacific. In addition, the model contains only one long memory predictor except for besides SST.

Among the three extra added physical variables, the upper ocean heat content is a very concerned variable, which can reflect the vertical and horizontal propagations of ocean waves and help interpret the dynamical mechanisms. Therefore, we conduct the comparison via two modified ENSO-ASCs with the same output of SST + u-wind, v-wind, rain, cloud, and vapor, while with the different input. One uses upper ocean heat content + u-wind, v-wind, rain, cloud, and vapor, marked as **EXAM**, another uses SST + u-wind, v-wind, rain, cloud, and vapor, marked as **CTRL**. The results are shown in Table 4.

Table 4: Model performance comparison when using upper ocean heat content to replace SST in the input

Model paradigm	12-month SSIM / PSNR	15-month SSIM / PSNR	18-month SSIM / PSNR
CTRL: SST + others → *SST + others*	92.65 / 22.05	90.31 / 20.97	87.53 / 18.17
EXAM: upper ocean heat content + others → *SST + others*	90.96 / 20.87	88.45 / 18.23	84.76 / 14.90

Note: Model paradigm represents the input and the output for the ENSO-ASC, where → means "forecast". "Others" represents five variables, including u-wind, v-wind, rain, cloud, and vapor. The first row is the control experiment, which is the same with the result in Table 3, and the second row is the examined experiment, which replaces SST by the upper ocean heat content in the model input.

The forecast skill of **EXAM** is slightly lower than **CTRL**. The upper ocean heat content is the average of the oceanic temperature from sea surface to upper 300m. When using it as a predictor to forecast SST, our model will extract the features of oceanic temperature not only from sea surface but also from deeper ocean, which inevitably introduces more noise. This may be a reason for the above result. Therefore, we still use SST instead of the upper ocean heat content as the key predictor which would bring higher forecast skills.

In the subsequent experiments, the model will use the chosen 6 variables (SST, u-wind, v-wind, rain, cloud, and vapor) and the input sequence length is set to 3.

In addition At the same time, we also add other variables into the multivariate coupler (surface air temperature/surface pressure/ocean heat content), then analyze the forecast skill change of the ENSO-ASC (the description of them are in the Table 3), the results exhibit little improvements a little improvement on the performance, shown as Table 3 (below). There are two three possible reasons for this result: Firstly, the contributions of the extra added variables to the evolution of ENSO is not as obvious as the existing chosen variables (such as pressure or air temperature), while the multivariate graph with existing variables can almost describe a relatively complete energy loop in Walker circulation. So the effect of extra variables for model is small. Secondly, the input sequence length for these experiments is also set as 3 months, which is beneficial for short

590 memory variables that account for most of the input variables, such as atmospheric variables (u-wind, v-wind, rain, cloud, and vapor). According to the previous analyses, the 3-month input length is an optimal determination due to less noise and high efficiency. However, this may be not suitable for long-memory variables, such as ocean heat content, which results in that the model cannot capture adequate long-term information from these variables and the improvement of forecast skill is insignificant. Thirdly, because the input region of our model majorly covers the equatorial Pacific but not the globe, the spatial
595 features of these variables are not fully utilized in the model as well. Take the heat content as an example, whichthe upper ocean heat content (vertically averaged oceanic temperature in the upper 300m) has more capabilities to describe the changes in the dynamic process of ENSO (Ham et al., 2019; McPhaden, 2003)., whichand has been formalized in the recharge oscillator theory of ENSO (Jin, 1997) and and has been verified against observations (Meinen and McPhaden, 2000). It can be reasonably speculated that the ocean heat content should promote the forecast skill more effectively. But due to the limitation of the input
600 region, the spatial teleconnections and temporal lagged correlations by Walker Circulation and ocean waves (such as Kelvin and Rossby Waves) (Exarchou et al., 2021 and Dommenget et al., 2006) cannot be fully considered and utilized in the current model, which hinders the further improvement of the forecast skill. Therefore, in the future study, we will carry out more comprehensive investigations on the effects of input sequence length and input variables fully according to the complex ENSO dynamical mechanisms using global data..

605 the concept of extra variables may be overlapped with the existing variables for the ENSO process. For example, the upper ocean heat content (vertically averaged oceanic temperature in the upper 300m) has more capabilities to describe the changes in the dynamic process of ENSO (Ham et al., 2019; McPhaden, 2003), which has been formalized in the recharge oscillator theory of ENSO (Jin, 1997) and has been verified against observations (Meinen and McPhaden, 2000). The calculation of this variable contains the region of SST, so the effect of the extra introduction of upper ocean heat content will
610 be weakened. In the subsequent experiments, we will still useFrom the perspective of dataset cost, the 6 chosen 6 variables (SST, u-wind, v-wind, rain, cloud, and vapor) in for the model, which can be obtained more easily with a quite high quality, and the model performance has reached an acceptable high level. While it can be reasonably speculated that the forecast skill of the coupler will be improved if considering more meteorological mechanisms with more crucial variables.

4.4.2 Analysis of Effective Forecast Lead Month

615 The accuracy of long-term prediction is the most crucial issue for meteorological research. In ENSO events, though the periodic dominants the amplitude, the intrinsic intensity and duration often contribute induce very large uncertainties and forecast errors. Therefore, over the validation period, we make predictions with multiple lead times and calculate the corresponding NiñoNiño indexes from the forecasted SST patterns to investigate the effective forecast lead month of our model. The correlations between forecasted NiñoNiño indexes and the official records are depicted in Fig. 14. As the lead time graduallys increases in
620 to 24 months, the correlation skill slowly decreases. It is worth noting that when the lead time is from 10 to 13 months, the decline of these three indexesthe forecast skillss slows down a little. This is because that the periodic in ENSO events becomes stronger after a 1-year iteration of in IMS strategy. These results demonstrate that the ENSO-ASC can provide reliable

25

predictions up to at least 18 to 20 months on average (with correlations over 0.5). Within 6-month lead time, the correlation skill is over about 0.78, and from 6- to 12-month lead time, correlation skill is over about 0.65.
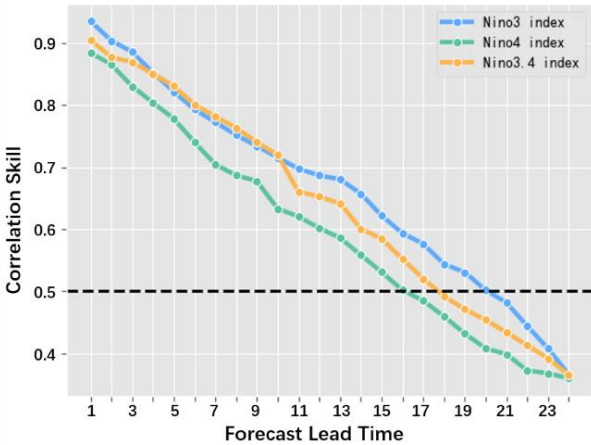

625

Figure 14: The correlation skills between the forecast results of the ENSO-ASC and real-world observations on three Niño indexes with the forecast lead time increasing over the validation period.

In addition, the forecast skills for ~~Niño~~Niño3 index and ~~Niño~~Niño3.4 index ~~is~~ are a little higher than that of ~~Niño~~Niño4 index. It indicates that our model has higher forecast skill for the EP-El ~~Niño~~Niño events (the active area of SST anomalies is majorly over the eastern tropical Pacific Ocean) than the CP-El ~~Niño~~Niño events (the active area of SST anomalies is majorly over the western and central tropical Pacific Ocean). This is because that the input area of the model majorly covers the entire tropical Pacific, which can be considered as the sensitive area for EP-El ~~Niño~~Niño events and is more favourable for the prediction of EP-El Niño events. As for ~~lower forecast skills~~the prediction of CP-El ~~Niño~~Niño events, extratropical Pacific or other oceanic regions may have stronger impacts on the western-central equatorial Pacific (Park et al., 2018).

4.4.3 Temporal Persistence Barrier with Different Start Calendar Months

Deep learning model can extend the effective lead time of ENSO forecasts, which means it can raise the upper limitation of ENSO prediction to some extent. From the perspective of IMS strategy, if a well-trained model can predict next-month SST perfectly (in other words, with a very low predict error), the model can iterate a lot~~forecast long enough iterations~~ theoretically. However, our proposed model is affected by a variety of factors, which leads to performance degradation.

One of the disadvantages in IMS strategy is that: Once a relatively large forecast error shows up in a certain iteration, such forecast error will be continuously amplified in subsequent iterations. In ENSO forecast, such forecast skill decline is ~~called~~ regarded as persistence barrier and usually ~~happens~~ occurs in spring (i.e. spring predictability barrier, SPB) (Webster, 1995; Zheng and Zhu, 2010). SPB limits the long-term forecast skill in not only numerical models but some other statistical models (Kirtman et al., 2001). For further investigating the ~~reason of~~ performance degradation, we firstly make continuous

645 predictions over the validation period from different start calendar months with different lead time, and then calculate the correlations between the calculated ~~Niño~~Niño3.4 index with the official records. Figure 15 shows the results.
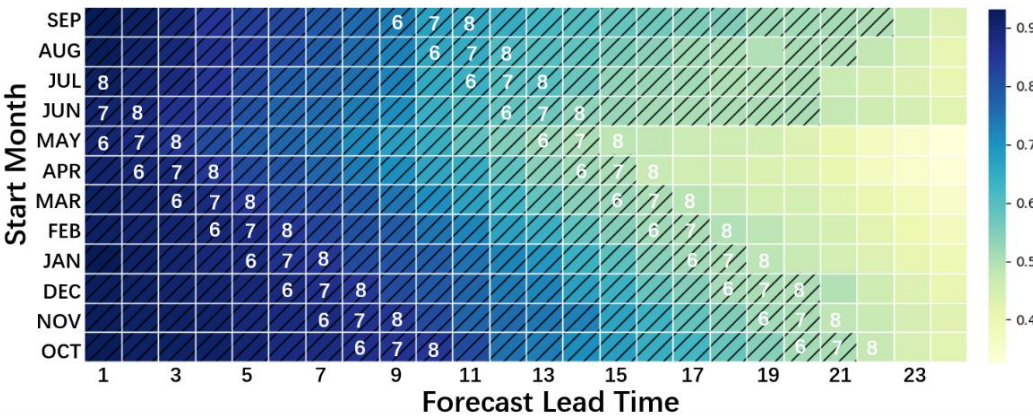
Figure 15: The correlation skill heat map between the forecast results of the ENSO-ASC and real-world observations on Niño3.4 index with different <u>forecast</u> start month<u>s</u> over the validation period. The hatching cells represent the correlation<u>s</u> ~~over~~ <u>exceed</u> 0.5,
650 **and the white numbers on the cells mean the ~~forecast~~ <u>calendar</u> months.**

In Fig. 15, the darker the cells' color, the higher the correlations between the forecast and observation, the higher the forecast skill. The hatching cells represent <u>that</u> the correlations ~~exceeding~~ <u>exceed</u> 0.5. Overall, making ESNO forecast with start calendar months MAM (March, April and May) is not very reliable, while the long-term forecast of ENSO is ~~easier~~ <u>more accurate</u> with start months JAS (July, August, and September). In addition, there exist two obvious color change zones among
655 the whole cells, which means the correlations drop~~ped~~ significantly in such zones (cell color becomes lighter), and both of them occur~~s~~ in the months JJA (June, July and August) depicted as the white numbers on the cells. The first zone reduces the correlations by about 0.03, and the second zone make<u>s</u> the reduction by about 0.06. It demonstrates that when making forecasts through the months JJA, the ENSO predictions tend to be much less successful. This is why the model exhibits more skilful forecast<u>s</u> with the start months JAS, which avoids forecasting through the months JJA as much as possible and preserves more
660 accurate features during iterations, resulting in a relatively long efficient lead time. Analogy to SPB in traditional ENSO forecast, our proposed ~~deep learning model~~ENSO-ASC exists forecast persistence barrier in boreal summer (JJA). This may be because that ~~our~~ <u>the</u> real-world dataset contains more frequency CP-El ~~Niño~~Niño ~~events~~ <u>samples</u> after 1990s (Kao and Yu, 2009; Kug et al., 2009), which ~~is~~ <u>are</u> significantly impacted by summer predictability barrier (Ren et al., 2019; Ren et al., 2016). At the same time, it also implies that there are still forecast obstacles ~~we~~ need to circumvent in the ~~deep learning~~ ENSO
665 <u>deep learning</u> forecast model, more unknown key factors need to be considered and explored, such as more variables, larger input regions, more complex mechanisms, etc. It will make great progress ~~from~~ <u>by</u> building deep learning model<u>s</u> based on prior meteorological knowledge <u>in the future</u>.

27

4.4.4 Spatial Uncertainties in Longer Lead Time

In ENSO forecast~~s~~, the areas where the forecast uncertainties occur are usually not randomly distributed, and such areas should

670 be ~~focused~~ paid more attention in operational target observation. Over the validation period, we make 12-month and 18-month

forecasts, and then compare the forecast results with observation~~s~~. More specifically, we calculate covariance between forecast

sequence $\hat{s}$ and observation s for every grid point and combine them as a spatial heat map.
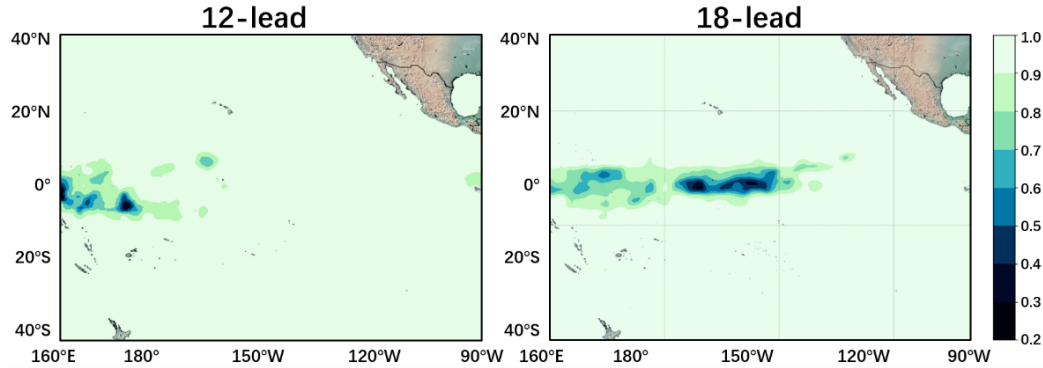
Figure 16: The spatial covariance heat map between the forecast results of the ENSO-ASC and real-world observations with 12 and
675 **18 lead month over the validation period.**

The results are shown in Fig. 16. The spatial uncertainties first show up over the western equatorial Pacific, then as the

lead time increases, the uncertainty area gradually expands eastward to the central equatorial Pacific. It indicates that the

~~Niño~~Niño3 and ~~Niño~~Niño3.4 ~~indexes~~ regions both have very high forecasting skills in a short forecast lead time (See Fig. 14

~~before~~ in 12-month forecast), while the ~~forecast~~ predictability for the central equatorial Pacific ~~is~~ gradually ~~disturbed~~ drops in

680 a longer lead time, which leads to a rapid reduction of forecast skill for ~~Niño~~Niño3.4 and ~~Niño~~Niño4 ~~indexes~~ regions shown

as ~~Fig. 14 after~~ 15-month forecast in Fig. 14. This ~~illustrates that the area with the largest forecast uncertainty is generally the~~

~~area with the most active SST evolution in ENSO~~reminds us that the areas with larger forecast uncertainties should be observed

by higher frequency. Besides this, another possible reason is that the ~~input~~ multivariate input region is confined to~~in~~ the Pacific,

but the ocean-atmosphere coupl~~ing~~ed interactions in the western tropical Pacific can be profoundly influenced by extratropical

685 Pacific areas ~~or~~ and other ocean basins as mentioned above. Therefore, our proposed model has relatively weak ability to

capture the development of SST over the western-central equatorial Pacific.

4.5 Simulation of the Real-world ENSO Events

Since the 21st century, the occurrences of ENSO are more and more frequent, especially the duration and intensity of ENSO

have largely changed. For example, many numerical climate models failed to forecast the 2015/2016 super El Niño. We

690 simulate several ENSO events during validation period and compare the forecast results with real-world observations. As

mentioned above, wind (u-wind and v-wind) is also a relatively important and sensitive predictor in the ENSO-ASC for ENSO

forecast~~s~~. Therefore, we make long-term forecasts and majorly trace the evolutions of SST and wind (u-wind and v-wind).
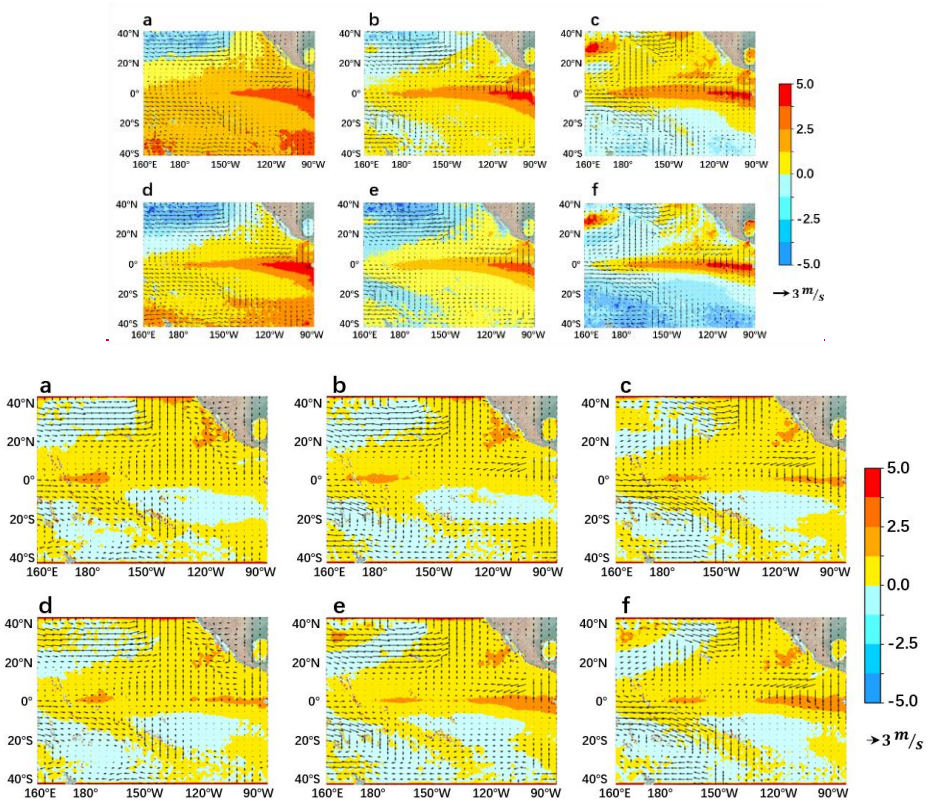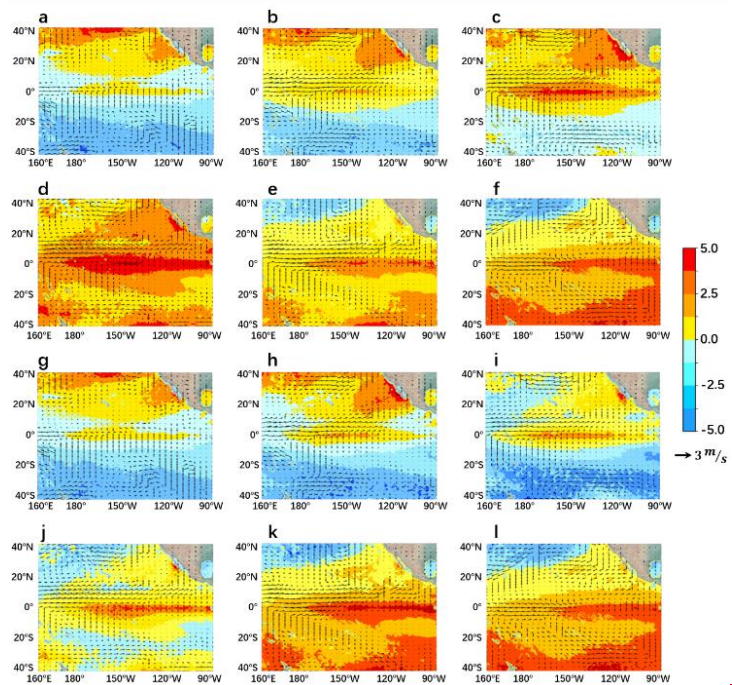
28

Figure 17: The growth phase of SST and wind anomalies in 2015/2016 super Niño event from 2015.4 to 2015.6. a-c are the forecast results of the ENSO-ASC and d-f are real-world observations.

Figure 17 displays the evolutions of SST and wind anomalies in growth phase of 2015/2016 super El Niño event from 2015.4 to 2015.6, where a-c sub-figures are forecasts and d-f sub-figures are observations. Fig. 18 displays the peak phase from 2015.9 to 2016.2, where a-f sub-figures are forecasts and g-l sub-figures are observations. These two results are both with the prediction forecast start time of 2015.1. During the growth phase, the warming SST anomalies first show up over the eastern tropical Pacific Ocean, which reduces the east-west gradient of SST. At the meanwhile, the westerly wind anomalies over the western-central equatorial Pacific further enhances the SST anomalies over central-eastern equatorial Pacific and weakens the Walker circulation (Fig. 17 a-c). The SST and wind anomalies trigger the Bjerknes positive feedback together, which causes SST anomalies to be continuously positively amplified. During the peak phase, in addition to the local evolutions of the equatorial Pacific SST anomalies, there are obvious warm SST anomalies over the northeast subtropical Pacific near

29

Baja California induced by the extratropical atmospheric varieties (Yu et al., 2010; Yu and Kim, 2011), which gradually propagates south-westward and merge with the warm SST anomalies over the central equatorial Pacific (Fig. 18 a-d). In conclusion, the ENSO-ASC can track the large-scale oceanic/atmospheric varieties steadily and can successfully predicts the ENSO with very highstrong -intensity and very-long -duration, while many dynamic or statistical models fail. At the same time, ourt proposed model makes the prediction at the beginning of the calendar year and produces a quite low prediction error, which demonstrates that the model can overcome or eliminate the negative impacts of SPB to some extent.
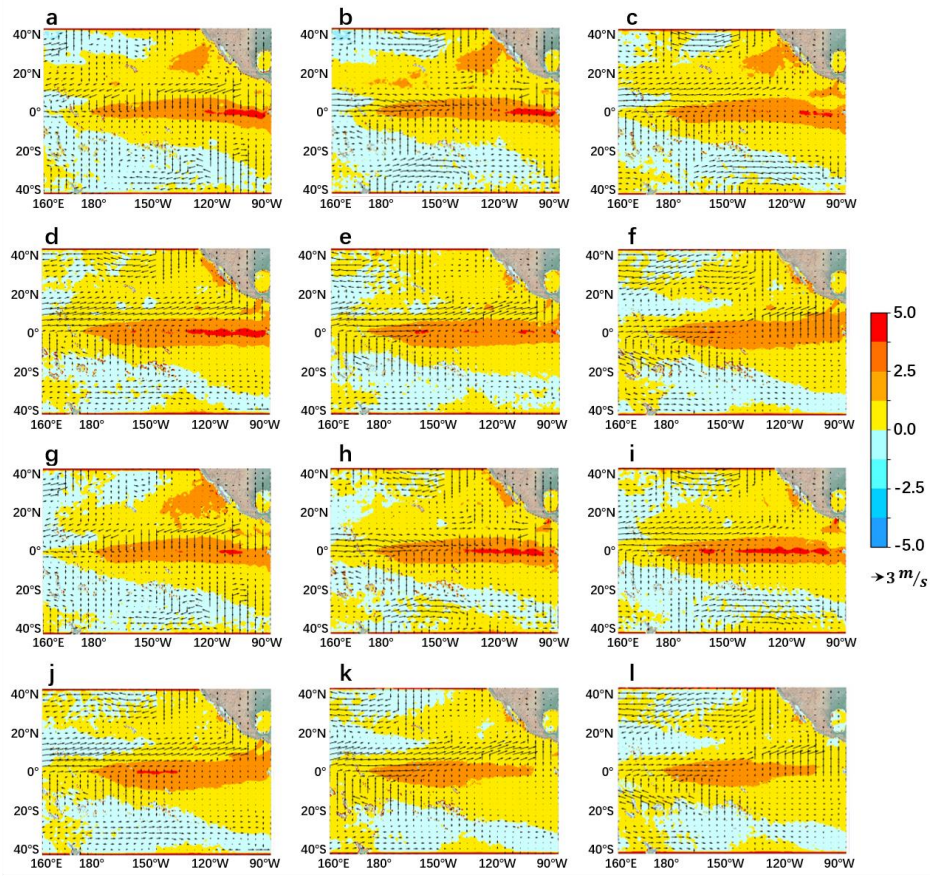
Figure 18~~18~~: The peak phase of SST _and wind_ anomalies in 2015/2016 super Niño event from 2015.9 to 2016.2. a-f are the forecast results of the ENSO-ASC and g-l are real-world observations.

~~From this real-world simulation, our model takes full advantage of multivariate coupled interactions and extract more nonlinear characteristics in the ENSO evolution, which has been tracking the developments steadily.~~
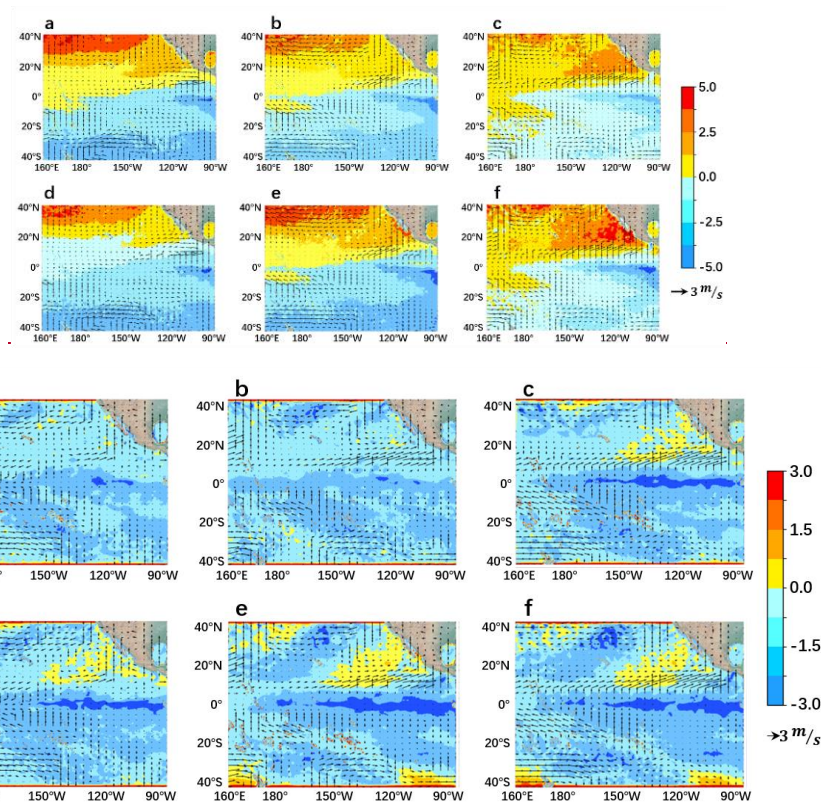
Figure 19: The same with Fig. 17 but for the growth phase of SST and wind anomalies of 2017 weak La Niña event from 2017.9 to 2017.11.

Besides the super El Niño event, the ENSO-ASC also has higher simulation capabilities for weaker nonlinear unstable evolutions. In reality, neutral or weaker events actually account for most of the time. Judging from the saliency of the extracted features, neutral/weaker events may contain "more mediocre and fuzzy" characteristics, which leads to some difficulties for such predictionsto grasp their meta features during evolutions accurately to some extent. For example, it is much easier to overestimate or underestimate their intensities. Therefore, we choose a hindcast over the validation period. Figure. 19 shows the peak phase of a weak La Niña event from 2017.9 to 2017.11 with the prediction forecast start time of 2016.9, where a-c sub-figures are forecasts and d-f sub-figures are observations. From its evolution, there are negative SST anomalies over the eastern equatorial Pacific and easterly wind anomalies in the western tropical Pacific Ocean, which will enhance the Walker circulation. In addition, Bjerknes positive feedback is the dominant factor favouring the rapid anomaly growth in this simulation.
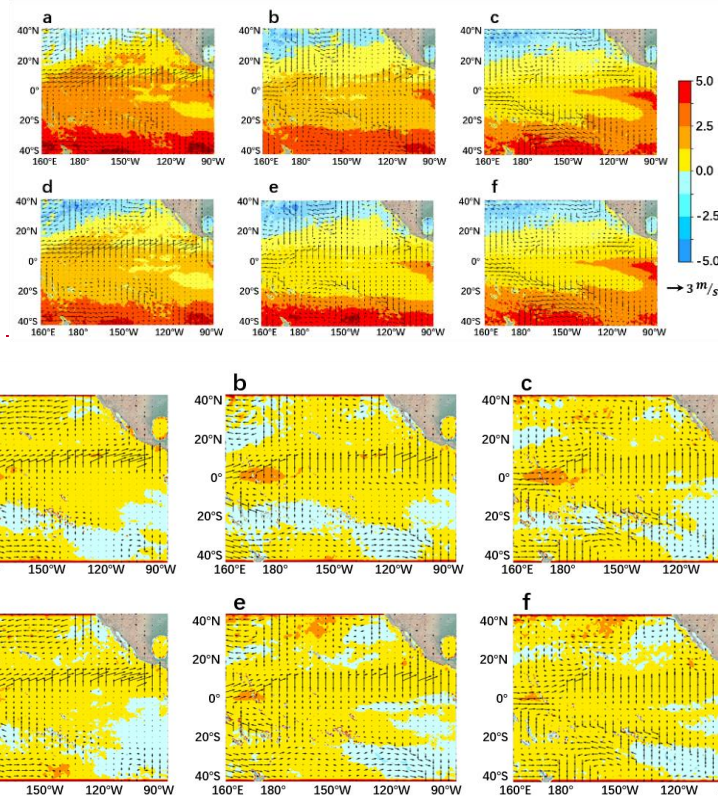
33

Figure 20: The same with Fig. 17 but for the neural SST and wind anomalies evolutions in 2020.1 to 2020.3.

Another ENSO forecast limitation is to predict the neural year as the event of El Niño (or La Niña), which is also known as false alarm rate. Figure 20 displays the neutral event from 2020.1 to 2020.3 with the ~~prediction~~ forecast start time of 2019.1 (Its ONI has not yet reached the intensity of El Niño). After calculating the corresponding Niño index, we can determine that the ENSO-ASC can accurately avoid the false alarm~~this situation~~ and credibly reflect the real magnitude of the development of important variables such as SST. We have also verified the case in 2014 and the result is consistent with the facts. Many operational centres erroneously predicted that an El Niño would develop in 2014, but it did not.

The forecasted SST and wind anomaly patterns have a consistent intensity and tendency with the observations. Why our model can achieve better forecast skills in a variety of situations? Because our proposed deep learning coupler comprehensively absorbs the sophisticated oceanic and atmospheric varieties, ~~whose~~ and its deep and intricate structure can almost simulate the air-sea energy exchange simultaneously, while traditional geoscience fluid programming in numerical climate model usually applies interval flux exchange and parameterized approximation for unknown mechanisms, blocking the continuous interactions.

34

5 Discussions and Conclusions

ENSO is a very complicated air-sea coupled phenomenon, the life-cycle of which is closely related to the large-scale nonlinear interactions between various oceanic and atmospheric processes. ENSO is one of the most critical factors that cause extreme climatic and socioeconomic effects. Therefore, meteorological researchers are starting up to find more accurate and less consuming data-driven models to forecast ENSO, especially deep learning methods. There are already many successful attempts that have extended the effective forecast lead time of ENSO up to one and a half years. They all extract the rich spatial-temporal information deeply hidden in the historical geoscience data.

However, most of the models use limited variables even single variable to predict ENSO, ignoring the coupling~~ed~~ multivariate interactions in ENSO events. At the same time, the generic ENSO deep learning forecast models seem to have reached the performance bottleneck, which means deeper or more complex model structures can neither extend the effective forecast lead time nor provide a more detailed description for dynamical evolutions. In order to overcome these two barriers, we subjectively incorporate priori ENSO knowledge into the deep learning formalization and derive hand-crafted features into models to make predictions. More specifically, considering the multivariate coupling in the Walker circulation related to ENSO amplitudes, we select 6 indispensable physical variables and focus on the synergies between them in ENSO events. Instead of simple variable stacking, we treat them as separate individuals and ingeniously formulate the nonlinear interactions between them on a graph. Based on such formalization, we design a multivariate air-sea coupler (ASC) by graph convolution mathematically, which can mine the coupling~~ed~~ interactions between every physical variable in pairs and perform the multivariate synergies simultaneously.

We then implement an ENSO deep learning forecast model (ENSO-ASC) with the encoder-coupler-decoder structure, and two self-supervised attention ~~components~~ weights are also ~~integrated~~designed. The multivariate time-series data is firstly ~~passed~~ propagated to the encoder to extract spatial-temporal features respectively.~~;~~ Then the multivariate features are aggregated together for interactions in the multivariate air-sea coupler.~~;~~ Finally, the coupled features are divided separately and the corresponding feature of a certain variable is restored to forecast patterns in the decoder respectively. IMS strategy is applied to make prediction~~s~~, which is a more stable forecast way. We use transfer learning to provide a better model initialization and overcome the problem of observation sample-lacking. The model is first trained on the reanalysis dataset and subsequently on remote sensing dataset. After constructing the model structure, we design extensive experiments to investigate the model performance and ENSO forecast skill. ~~Some~~ Several successful simulations in the validation period are also provided. Some conclusions can be summarized as follows:

1) According to the forecast model described in Eq. (5) to Eq. (7), we adjust the model settings of the input sequence length M, the multivariate coupler $coupler(\cdot)$, the attention weights α and β, the transfer training, and then investigate the performance change~~s~~. The optimal input sequence length of the model is 3 according to the trade-off between forecast skill and ~~speed~~ computational resource consuming. It implies that the ENSO deep learning forecast model does not need a relatively long-sequence input. Although the long sequence contains rich tendency and periodic of ENSO events, the

785   meteorological chaos is more dominant, which seriously hinders the prediction. Transfer learning is a practical way. Training the model on the reanalysis dataset and subsequently on the remote sensing dataset can effectively reduce the systematic forecast errors by at least 15%. When replacing the graph-based multivariate air-sea coupler in ENSO forecast model with other deep learning structures, the forecast skill drops obviously. This demonstrates that the graph formalization is a powerful expression for simulating underlying air-sea interactions and corresponding ENSO forecast

790   model with novel multivariate air-sea coupler can forecast more realistic meteorological details. This also demonstrates that it is critical to choose suitable deep learning structures to incorporate prior climate mechanisms for improving forecast skills. The self-supervised attention ~~components~~ weights are also promising tools to ~~eliminate the performance reductions caused by~~grasp the contributions of different ~~physical variables~~predictors and memory varieties of different forecast start calendar month. In addition~~.~~ ~~by~~comparing with other state-of-the-art ENSO forecast models, the ENSO-

795   ASC achieves at least 5% improvement in *SSIM* and *PSNR* ~~of~~for long-term forecast~~ed SST pattern~~s.

2) By performing the ablation experiment, the forecast skill drops significantly when removing the zonal wind ~~speed~~from the model input, which is because it is a co-trigger of Bjerknes positive feedback in ENSO events and gives a direct feedback on oceanic varieties with a shorter lag time. Adding extra ~~variables~~predictor can slightly improve the performance, ~~especially the ocean temperature heat content, which is the concept and calculation extension of the original~~

800   ~~SST~~this is because that the existing multivariate graph can almost describe a relatively complete energy loop in Walker circulation. By tracing the upper limitation of forecast lead time, the ENSO-ASC can provide a reliable high-resolution ENSO forecast up to at least 18 to 20 months on average judging from the correlation skills of Niño indexes greater than 0.5. Within 6-month lead time, the correlation skill is over about 0.78, and from 6- to 12-month lead time, correlation skill is over about 0.65. The corresponding correlation skills decline slowly from 10- to 13-month lead time, and then

805   declined rapidly. This is because of the stronger periodic in ENSO events after a 1-year iteration of IMS strategy. At the same time, the different forecast start calendar months also ~~is~~influen~~cetial for~~ the ~~performance~~forecast skills. The temporal heat map analysis shows that an obvious skill reduction usually shows up in JJA and produces a boreal summer persistence barrier in our model. In addition, from the spatial uncertainty heat map, our model exhibits larger forecast uncertainties over the ~~SST active region, especially~~ western-central equatorial Pacific. Such spatial-temporal

810   predictability barriers are widely present in dynamic or statistical models, but the ENSO-ASC effectively prolongs the forecast lead time and reduces corresponding uncertainties to a large extent.

3) Some successful simulations exhibit the effectiveness and superiority of the ENSO-ASC. We make real-world ENSO simulations during the validation period by tracing the evolutions of SST and wind anomalies (u-wind and v-wind). In the forecasted El Niño (La Niña) events, the sea-air patterns clearly display that the positive (negative) SST anomalies

815   first show up over the eastern equatorial Pacific with westerly (easterly) wind anomalies in the western-central tropical Pacific Ocean, which induces the Bjerknes positive feedback mechanism. As for 2015/2016 super El Niño, the ENSO-ASC captures the strong evolutions of SST anomalies over the northeast subtropical Pacific in the peak phase and successfully predicts its very-high-intensity and very-long-duration, while many dynamic or statistical models fail.

36

ENSO-ASC can also credibly reflect the real situation and reduce the false alarm rate of ENSO such as in 2014. In
820   conclusion, our model can track the large-scale oceanic and atmospheric varieties and simulate the air-sea energy
exchange simultaneously. It demonstrates that the multivariate air-sea coupler effectively simulates the oscillations of
Walker circulation and reveals more complex dynamic mechanisms such as Bjerknes positive feedback.

The extensive experiments demonstrate that the ENSO forecast model with a multivariate air-sea coupler (ENSO-ASC)
is a powerful tool for analysis of ENSO-related complex mechanisms. Meteorological research does not only pursue skilful
825   models and accurate forecasts, but requires a comprehensive understanding of the potential dynamical mechanisms.~ In the
future, we will extend our model to more global physical variables with informative vertical layers, such as the thermocline
depth, and the ocean temperature heat content, ~~The extensive experiments demonstrate that the ENSO forecast model with a~~
~~multivariate air-sea coupler (ENSO-ASC) is a powerful tool for analysis of ENSO-related complex mechanisms~~to explore the
global spatial remote teleconnections, temporal lagged correlations, and the optimal precursor etc.~~.~~

830   ~~. Thus we will use it in the ENSO predictability research in depth. Firstly, we will extend our model to longer input~~
~~sequence length with global data for more comprehensive representations for SST varieties in ENSO, considering the global~~
~~spatial remote teleconnections and temporal lagged correlations. Then w~~We will identify the optimal precursor by involving
~~more physical variables with informative vertical layers (such as the thermocline depth, and the ocean temperature heat content)~~
~~considering more sophisticated dynamic mechanisms. Furthermore, we will carry out sensitivity ENSO mechanism~~
835   ~~investigation via our model, revealing the most significant temporal correlations and spatial uncertainties within multiple~~
~~ENSO related variables, which can promote interdisciplinary research between computer science and meteorological~~
~~geoscience.~~

Code availability

The source code of the ENSO-ASC is available in the Git repository: https://github.com/BrunoQin/ENSO-ASC (last access:
840   14 August 2021), which is implemented by Python 3.6 (or 3.7) and CUDA 11.0. The present version of ENSO-ASC 1.0.0 is
available at https://doi.org/10.5281/zenodo.5081793.~~The source code of the ENSO-ASC 1.0.0 is available at~~
~~https://doi.org/10.5281/zenodo.5081794.~~

Data availability

Thanks to NOAA/CIRES, Remote Sensing System, and China Meteorological Administration for providing the historical
845   geoscience data and analysis tools. (https://rda.ucar.edu/, http://www.remss.com/, https://cmdp.ncc-cma.net, last access: 8 Jul
2021). The related training/validation datasets can be also accessed at https://doi.org/10.5281/zenodo.5179867~~Thanks to~~
~~NOAA/CIRES, Remote Sensing System, and China Meteorological Administration for providing the historical geoscience~~
~~data and analysis tools. (https://rda.ucar.edu/, http://www.remss.com/, https://cmdp.ncc-cma.net, last access: 8 Jul 2021).~~

Author contribution

850 All authors design the experiments carry them out. Bo Qin develops the model code and performs the simulations. Bo Qin and Shijin Yuan prepare the manuscript with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgment

References

860 Balmaseda, M. A., Davey, M. K., and Anderson, D. L.: Decadal and seasonal dependence of ENSO prediction skill, Journal of Climate, 8, 2705–2715, 1995.

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., and 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, 2012.

865 Bayr, T., Dommenget, D., and Latif, M.: Walker circulation controls ENSO atmospheric feedbacks in uncoupled and coupled climate model simulations, Climate Dynamics, pp. 1–16, 2020.

Bellenger, H., Guilyardi, É., Leloup, J., Lengaigne, M., and Vialard, J.: ENSO representation in climate models: From CMIP3 to CMIP5, Climate Dynamics, 42, 1999–2018, 2014.

Bjerknes, J.: Atmospheric teleconnections from the equatorial Pacific, Monthly weather review, 97, 163–172, 1969.

870 Broni-Bedaiko, C., Katsriku, F. A., Unemi, T., Atsumi, M., Abdulai, J.-D., Shinomiya, N., and Owusu, E.: El Niño-Southern Oscillation forecasting using complex networks analysis of LSTM neural networks, Artificial Life and Robotics, 24, 445–451, 2019.

Bruna, J., Zaremba, W., Szlam, A., and LeCun, Y.: Spectral networks and locally connected networks on graphs, arXiv preprint arXiv:1312.6203, 2013.

875 Chen, F., Pan, S., Jiang, J., Huo, H., and Long, G.: DAGCN: Dual Attention Graph Convolutional Networks, 2019.

Cheng, L., Trenberth, K. E., Fasullo, J. T., Mayer, M., Balmaseda, M., and Zhu, J.: Evolution of ocean heat content related to ENSO, Journal of Climate, 32, 3529–3556, 2019.

Chevillon, G.: Direct multi-step estimation and forecasting, Journal of Economic Surveys, 21, 746–785, 2007.

Defferrard, M., Bresson, X., and Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering, arXiv preprint arXiv:1606.09375, 2016.

Dommenget, D., Semenov, V., and Latif, M.: Impacts of the tropical Indian and Atlantic Oceans on ENSO, Geophysical research letters, 33, 2006.

Exarchou, E., Ortega, P., Rodríguez-Fonseca, B., Losada, T., Polo, I., and Prodhomme, C.: Impact of equatorial Atlantic variability on ENSO predictive skill, Nature communications, 12, 1–8, 2021.

Gao, C. and Zhang, R.-H.: The roles of atmospheric wind and entrained water temperature (T e) in the second-year cooling of the 2010–12 La Niña event, Climate Dynamics, 48, 597–617, 2017.

Ham, Y.-G., Kim, J.-H., and Luo, J.-J.: Deep learning for multi-year ENSO forecasts, Nature, 573, 568–572, 2019.

Hammond, D. K., Vandergheynst, P., and Gribonval, R.:Wavelets on graphs via spectral graph theory, Applied and Computational Harmonic Analysis, 30, 129–150, 2011.

He, D., Lin, P., Liu, H., Ding, L., and Jiang, J.: Dlenso: A deep learning enso forecasting model, in: Pacific Rim International Conference on Artificial Intelligence, pp. 12–23, Springer, 2019.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q.: Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708, 2017.

Jin, F.-F.: An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model, Journal of the atmospheric sciences, 54, 811–829, 1997.

Kao, H.-Y. and Yu, J.-Y.: Contrasting eastern-Pacific and central-Pacific types of ENSO, Journal of Climate, 22, 615–632, 2009.

Keys, R.: Cubic convolution interpolation for digital image processing, IEEE transactions on acoustics, speech, and signal processing, 29, 1153–1160, 1981.

Kirtman, B., Shukla, J., Balmaseda, M., Graham, N., Penland, C., Xue, Y., and Zebiak, S.: Current status of ENSO forecast skill: A report to the CLIVAR Working Group on Seasonal to Interannual Prediction, 2001.

Kug, J.-S., Jin, F.-F., and An, S.-I.: Two types of El Niño events: cold tongue El Niño and warm pool El Niño, Journal of Climate, 22, 1499–1515, 2009.

Lau, K.-M., Li, P., and Nakazawa, T.: Dynamics of super cloud clusters, westerly wind bursts, 30-60 day oscillations and ENSO: An unified view, Journal of the Meteorological Society of Japan. Ser. II, 67, 205–219, 1989.

Lau, K.-M., Ho, C.-H., and Chou, M.-D.: Water vapor and cloud feedback over the tropical oceans: Can we use ENSO as a surrogate for climate change?, Geophysical research letters, 23, 2971–2974, 1996.

910   Mazumder, R., Hastie, T., and Tibshirani, R.: Spectral regularization algorithms for learning large incomplete matrices, The Journal of Machine Learning Research, 11, 2287–2322, 2010.

McDermott, P. L. and Wikle, C. K.: An ensemble quadratic echo state network for non-linear spatio-temporal forecasting, Stat, 6, 315–330, 2017.

McDermott, P. L. andWikle, C. K.: Bayesian recurrent neural network models for forecasting and quantifying uncertainty in
915   spatial-temporal data, Entropy, 21, 184, 2019.

McPhaden, M. J.: Tropical Pacific Ocean heat content variations and ENSO persistence barriers, Geophysical research letters, 30, 2003.

McPhaden, M. J.: A 21st century shift in the relationship between ENSO SST and warm water volume anomalies, Geophysical research letters, 39, 2012.

920   Meinen, C. S. and McPhaden, M. J.: Observations of warm water volume changes in the equatorial Pacific and their relationship to El Ni ño and La Ni ña, Journal of Climate, 13, 3551–3559, 2000.

Mu, B., Peng, C., Yuan, S., and Chen, L.: ENSO forecasting over multiple time horizons using ConvLSTM network and rolling mechanism, in: 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, IEEE, 2019.

Park, J.-H., Kug, J.-S., Li, T., and Behera, S. K.: Predicting El Ni ño beyond 1-year lead: effect of the Western Hemisphere
925   warm pool, Scientific reports, 8, 1–8, 2018.

Ren, H.-L., Jin, F.-F., Tian, B., and Scaife, A. A.: Distinct persistence barriers in two types of ENSO, Geophysical Research Letters, 43, 10–973, 2016.

Ren, H.-L., Zuo, J., and Deng, Y.: Statistical predictability of Ni ño indices for two types of ENSO, Climate Dynamics, 52, 5361–5382, 2019.

930   Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., et al.: Tackling climate change with machine learning, arXiv preprint arXiv:1906.05433, 2019.

Shi, X. and Yeung, D.-Y.: Machine learning for spatiotemporal sequence forecasting: A survey, arXiv preprint arXiv:1808.06865, 2018.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c.: Convolutional LSTM network: A machine learning
935   approach for precipitation nowcasting, arXiv preprint arXiv:1506.04214, 2015.

Wang, C., Deser, C., Yu, J.-Y., DiNezio, P., and Clement, A.: El Ni ño and southern oscillation (ENSO): a review, Coral reefs of the eastern tropical Pacific, pp. 85–106, 2017.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image quality assessment: from error visibility to structural similarity, IEEE transactions on image processing, 13, 600–612, 2004.

940   Webster, P.: The annual cycle and the predictability of the tropical coupled ocean-atmosphere system, Meteorology and Atmospheric Physics, 56, 33–55, 1995.

Xue, Y., Chen, M., Kumar, A., Hu, Z.-Z., and Wang, W.: Prediction skill and bias of tropical Pacific sea surface temperatures in the NCEP Climate Forecast System version 2, Journal of climate, 26, 5358–5378, 2013.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H.: How transferable are features in deep neural networks?, arXiv preprint
945   arXiv:1411.1792, 2014.

Yu, J.-Y. and Kim, S. T.: Relationships between extratropical sea level pressure variations and the central Pacific and eastern
        Pacific types of ENSO, Journal of Climate, 24, 708–720, 2011.

Yu, J.-Y., Kao, H.-Y., and Lee, T.: Subtropics-related interannual sea surface temperature variability in the central equatorial
        Pacific, Journal of Climate, 23, 2869–2884, 2010.

950   Zhang, R., Liu, Q., and Hang, R.: Tropical Cyclone Intensity Estimation Using Two-Branch Convolutional Neural Network
        From Infrared and Water Vapor Images, IEEE Transactions on Geoscience and Remote Sensing, 58, 586–597, 2019a.

Zhang, W., Jin, F.-F., Stuecker, M. F., Wittenberg, A. T., Timmermann, A., Ren, H.-L., Kug, J.-S., Cai, W., and Cane, M.:
        Unraveling El Niño's impact on the East Asian monsoon and Yangtze River summer flooding, Geophysical Research
        Letters, 43, 11–375, 2016.

955   Zhang, W., Li, S., Jin, F.-F., Xie, R., Liu, C., Stuecker, M. F., and Xue, A.: ENSO regime changes responsible for decadal
        phase relationship variations between ENSO sea surface temperature and warm water volume, Geophysical Research
        Letters, 46, 7546–7553, 2019b.

Zheng, F. and Zhu, J.: Spring predictability barrier of ENSO events from the perspective of an ensemble prediction system,
        Global and Planetary Change, 72, 108–117, 2010.

960   Zheng, G., Li, X., Zhang, R.-H., and Liu, B.: Purely satellite data–driven deep learning forecast of complicated tropical
        instability waves, Science advances, 6, eaba1482, 2020.