imts Intelligent Marine
Technology and Systems

RESEARCH PAPER

Open Access

Multi-scale spatial features and temporal attention mechanisms: advancing the accuracy of ENSO prediction

Shengen Tao[2], Yanqiu Li[2], Feng Gao[1], Hao Fan[1], Junyu Dong[1,2*] and Yanhai Gan[1*]

Abstract

The exponential progression in oceanic observational technology has fostered the accumulation of substantial time series data pivotal for predictions in ocean meteorology. Foremost among the phenomena observed is El Niño-Southern Oscillation (ENSO), a critical determinant in the interplay of global ocean atmosphere interactions, with its severe manifestations inducing extreme meteorological conditions. Therefore, precisely predicting ENSO events carries immense gravitas. Historically, predictions hinged primarily on dynamic models and statistical approaches; however, the intricate and multifaceted spatiotemporal dynamics of ENSO events have often impeded the accuracy of these traditional methodologies. A notable lacuna in contemporary research is the insufficient exploration of long-term dependencies within oceanic data and the suboptimal integration of spatial information derived from spatiotemporal data. To address these limitations, this study introduces a forward-thinking ENSO prediction framework synergizing multiscale spatial features with temporal attention mechanisms. This innovation facilitates a more profound exploration of temporal and spatial domains, enhancing the retention of extensive-period data while optimizing the use of spatial information. Preliminary analyses executed on the global ocean data assimilation system dataset attest to the superior efficacy of the proposed method, underscoring a substantial improvement over established methods including SA-convolutional long short-term memory, particularly in facilitating long-term predictions. The source code and datasets are provided. The code is available at https://github.com/tse1998/ENSO-prediction.

Keywords Deep learning, Multiscale spatiotemporal features, ENSO prediction, Sequence prediction

1 Introduction

Many ocean phenomena have a profound connection to societal well-being, with El Niño-Southern Oscillation (ENSO) being a principal determinant of climate fluctuations globally. This phenomenon garners substantial attention because of its role in engendering natural disasters worldwide. The ENSO phenomenon manifests through variations in sea surface temperature (SST) in the tropical Pacific, elevating or reducing it beyond the average value, consequently affecting global rainfall patterns and initiating a series of climatic issues worldwide. Research indicates that ENSO monumentally impacts the global economy. In the United States alone, extreme weather events attributed to ENSO have incurred losses amounting to billions of dollars (Adams et al. 1999). Furthermore, the repercussions of ENSO are not confined to economic dimensions; it also considerably influences the trajectory of social development (Glantz 2001). Thus, the accurate prediction of extreme ENSO events is a crucial objective, the fulfillment of which could bestow widespread benefits on society.

*Correspondence:
Junyu Dong
dongjunyu@ouc.edu.cn
Yanhai Gan
ganyanhai@ouc.edu.cn
[1] Faculty of Information Science and Engineering, Ocean University of China, Qingdao 266100, China
[2] Sanya Oceanographic Institution, Ocean University of China, Sanya 572000, China

Meteorological predictions used in oceanography predominantly adhere to two methodologies: numerical meteorological predictions and data-driven approaches. Numerical predictions rely heavily on meteorological dynamics, and physical models grounded in meteorological principles are constructed to facilitate weather forecasts. This strategy, which has evolved to a mature state with sustained development, involves solving intricate physical models to derive forecast outcomes, thus standing as a vital tool in weather forecasting. Conversely, data-driven approaches employ statistical or machine learning algorithms to anticipate future meteorological trends, leveraging extensive data repositories for prediction. Although numerical forecasting methodologies are considerably mature, they grapple with limitations such as high computational costs and sluggish response times, which are due to the complexities of synchronizing physical models and the labor-intensive nature of expansive numerical simulations. In the context of ENSO prediction, the prevailing numerical prediction paradigms exhibit deficiencies in effectively simulating average annual variations in SST, which dampens their predictive efficacy (Jin et al. 2008). The ENSO phenomenon, characterized by notable spatiotemporal variability and diversity, poses an intrinsic challenge to these traditional forecasting techniques, resulting in a pronounced level of uncertainty in predictions (Jiang et al. 2016).

Recently, advances in big data technology have facilitated an exponential increase in the accumulation of observational data in the realm of ocean science, with a considerable amount of this information being spatiotemporal data amenable to meteorological predictive analyses. This burgeoning data reservoir has underscored the potential of data-driven methodologies, propelling them to the forefront of research avenues in this field. To address the shortcomings inherent in numerical meteorological forecasting, researchers are increasingly leveraging machine learning and deep learning techniques synergized with ocean data to enhance meteorological predictions. These techniques encompass a spectrum of algorithms, such as support vector machines (SVMs), convolutional neural networks (CNNs), and convolutional long short-term memory (ConvLSTM) networks (Aguilar-Martinez and Hsieh 2009; Ham et al. 2019; He et al. 2019). The focal objective underpinning the use of these advanced algorithms is to delineate nonlinear mappings of pertinent features within the oceanic data spectrum. However, ENSO prediction fundamentally remains a sequential prediction issue, necessitating the integration of historical data-including past SST data-to anticipate future meteorological alterations. Despite the strides made, the high dimensionality of spatiotemporal data introduces a complexity warranting further optimization of the existing methods to fully exploit the spatiotemporal information pertinent to ENSO data for more refined predictive outcomes.

Despite concerted efforts from the global research community to advance marine hydrometeorological forecasting, and notwithstanding the progress achieved in ocean data prediction, existing methodologies continue to grapple with two central unresolved challenges:

(1) Suboptimal exploration of long-term dependencies: The prevailing focus of contemporary research predominantly revolves around modeling short-term relationships within sequential data, neglecting to unearth and leverage the potential insights derivable from long-term dependencies in the data.
(2) Limited exploitation of spatial information: Marine phenomena unfold through intricate spatiotemporal processes encompassing fluctuations across time and complex interactions in the spatial dimension. However, current methodologies only superficially tap into the rich tapestry of spatial information, forfeiting the depth of analysis that could enhance forecasting accuracy and insight.

To address the aforementioned challenges, the main contributions of this study are as follows:

(1) The introduction of a temporal attention module into the seq2seq framework marks a pivotal development. This module facilitates the intricate modeling of time series data along the temporal axis, enabling the discernment of evolving patterns and trends pertinent to ENSO phenomena. Furthermore, it leverages an attention mechanism in the time series prediction, fostering adaptive learning and precise identification of critical temporal nodes, thereby augmenting the accuracy of ENSO predictions over extended timeframes.
(2) The inception of a feature extraction module grounded in multi-scale convolution represents an important step. This initiative heralds the integration of multi-scale convolution techniques into spatiotemporal data processing, which is pivotal for assimilating spatial information from ocean meteorology. Consequently, it engenders a rich tapestry of spatial features discerned at various scales, enriching the repository of spatial dimension information and enhancing analytical depth.
(3) The proposition of an ENSO prediction model orchestrated around multi-scale spatial features and temporal attention is a cornerstone contribution. The judicious combination of the multi-scale convolution and temporal attention module fosters

a holistic exploration of the input sequence data across temporal and spatial dimensions. This integrated approach, consequently, holds substantial promise in augmenting the predictive accuracy and performance of ENSO forecasting models.

This study introduces an ENSO prediction model grounded in multi-scale spatial features and temporal attention mechanisms. Leveraging the temporal attention framework facilitates a more robust extraction of long-term dependencies within ocean data, enhancing the model's ability to simulate dynamics over extended periods. Concurrently, multi-scale convolution techniques are adopted to amplify the extraction of spatial information and multi-scale features from spatiotemporal data, overcoming the existing limitations in spatial data use. To substantiate the efficacy of the developed model, a series of tests were executed using the global ocean data assimilation system (GODAS) dataset as a benchmark. The preliminary results indicate a superior performance of the proposed model relative to established machine learning and deep learning approaches, demonstrating enhanced predictive accuracy over a 24-month forecast horizon.

2 Related work

Research into the prediction of ENSO primarily relies on two preeminent methodologies: numerical weather prediction and data-driven approaches. The former methodology leverages the foundational principles of ocean-atmosphere dynamics inherent in marine science to craft physical models that delineate the interrelations between atmospheric and oceanic variables, constituting a pivotal tool in ENSO forecasting. Conversely, data-driven strategies employ statistical analyses coupled with machine learning or deep learning techniques to exploit extant data pools to anticipate forthcoming meteorological shifts. These methodologies offer a heightened level of accuracy and applicability in short- to medium-term forecasts when juxtaposed with their model-based counterparts. This superiority stems from the intrinsic capacity of data-driven approaches to thoroughly elucidate patterns and trajectories within expansive historical datasets, facilitating adaptive learning and predictions of emergent data (Barnston et al. 2012). Meanwhile, model-based techniques remain constrained by the confines of model precision and the depth of understanding of the ocean-atmosphere system, which frequently hampers the attainment of accurate prognostications regarding future meteorological alterations. In contemporary meteorological and oceanographic research spheres, data-driven methods are consequently witnessing escalating adoption for their prowess in forecasting meteorological and

oceanic phenomena, solidifying their role as indispensable tools in the modern predictive landscape.

(1) Numerical weather forecasting

The symbiotic relationship between atmospheric and oceanic phenomena plays a pivotal role in steering long-term global climate transitions, with ENSO epitomizing a crucial aspect of this complex interplay. Predicated on a coupled oscillation involving the tropical Pacific Ocean and the atmosphere, the ENSO system stands as a beacon in understanding the synergistic ocean-atmosphere dynamics. In delving into ENSO forecasting, researchers have traversed a gamut of approaches grounded in ocean dynamics (Chen et al. 2004), ranging from rudimentary tropical ocean-atmosphere coupling models to intricately designed models (Zhang 2015; Zheng and Zhu 2016) and multifaceted global models encapsulating a broader spectrum of ocean-atmosphere interactions (Luo et al. 2008). These endeavors seek a foundation in a plethora of theories, such as the equatorial wind stress hypothesis, coupled ocean-atmosphere wave theory in the tropics, and the delayed oscillator mechanism, each offering a distinct lens for comprehending the underlying physical processes. Despite these sustained efforts in dynamically modeled predictive methodologies, the scientific community has yet to conceive a comprehensive theoretical framework encapsulating the complete El Niño system. This shortfall underscores the prevalent challenges and perhaps limitations in leveraging atmospheric and marine dynamics theories for ENSO prediction, indicating a substantial scope for further refinement and exploration in this field.

(2) Data-driven methods

Data-driven approaches involve applying statistical and computational learning techniques to delineate the intricate relationships between predictive elements and the ENSO index. The methodologies within this realm span a diverse range, including persistence forecasting, similarity analyses, and a spectrum of linear models such as multivariate regression and Markov chains, along with canonical correlation analyses. These methods manifest in three distinct categories. The inaugural category embraces statistical theory-grounded models such as autoregressive integrated moving average and kernel function statistical models, offering a mathematical framework for prediction. Following this strategy, traditional machine learning algorithms and their derivatives form the second category, encompassing techniques such as Bayesian neural networks, support vector regression, and SVM that facilitate nuanced analyses.

The final category represents the frontier of deep learning techniques, leveraging architectures such as LSTM networks, ConvLSTM, and transformers, which have heralded promising results in identifying complex data patterns unrecognizable through conventional statistical paradigms (Mu et al. 2019). These avant-garde methodologies have not only advanced ENSO predictive modeling but also found pivotal applications in diverse sectors, including meteorological forecasting, natural disaster anticipatory models, and the financial analytics domain, showcasing their extensive utility and potent efficacy in data analysis.

3 Method

3.1 seq2seq framework

In the ever-evolving field of time series forecasting, methods such as recurrent neural networks, LSTM, and ConvLSTM have generally outperformed traditional approaches, particularly in the context of sequence predictions. These techniques have shown superior capabilities in uncovering underlying patterns or trends in the input sequence data. Nevertheless, current predictive models predominantly focus on analyzing data through the temporal dimension, leaving considerable room for enhancing the depth of exploration into the complex trends in data sequences. This situation underscores the imperative for improved accuracy and the development of methodologies capable of long-term predictions, which are essential in ENSO forecasting.

To address this gap, we introduce an advanced ENSO predictive model grounded in the temporal attention mechanism and using the seq2seq framework. This structured approach facilitates a more thorough analysis of the input sequence data, effectively navigating the challenges presented by long-term dependencies. Distinctively, the devised generative model can accommodate variable-length sequence alignments, capitalizing on the correlations between output labels to undertake multistep forward predictions. As depicted in Fig. 1, this strategy offers a comprehensive solution, promising enhanced precision and extended foresight in sequence predictions, thus meeting the complex requirements of ENSO forecasting.

The seq2seq framework leverages the LSTM architecture as its foundation. This architecture is structured into two primary components: an encoder and a decoder. The encoder accepts sequence data in increments of a 12-step size as input, covering a geographical expanse from a southern latitude of 55° to a northern latitude of 60° and an eastern longitude scope from 0° to 360°. These data encompass variables such as SST and ocean heat content anomaly data, represented at each time step t with a dimensional scope denoted by $C \times H \times W$. Within the encoder, a tri-layer LSTM infrastructure operates, where each time increment is distilled into a one-dimensional vector before being channeled into the LSTM. Here, the innovative gating mechanism of the LSTM comes into play, diligently learning and adapting to the input data.
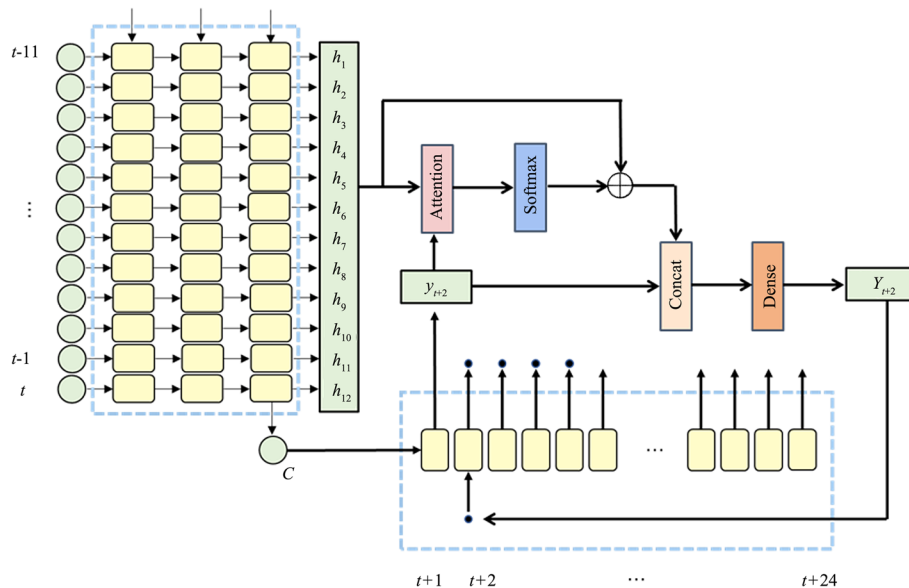
Fig. 1 Schematic of the temporal attention mechanism: the encoder-decoder and the temporal attention mechanism are parts of the model framework, with a three-layer LSTM architecture on the encoder side and a one-layer LSTM architecture on the decoder side. In this figure, h is the state vector of all encoders, v_t is the output of the decoder, Y_t is the prediction result of the current model and is used as the input data for the next moment, and C is the output of the encoder, representing the entire input time series data, which go into the decoder for processing

The decisive role of the forget gate in this structure cannot be overstated, as it judiciously determines the volume of information to be retained from the preceding time step $t-1$, ensuring a responsive and intelligent analysis grounded in pertinent historical data.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \tag{1}$$

The input gate determines which information should be retained at time t.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{2}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{3}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t. \tag{4}$$

Subsequently, the output gate discerns the data to be transmitted to the output stream h_t and the data destined for the memory stream C_t.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{5}$$

$$h_t = o_t \cdot \tanh(C_t). \tag{6}$$

Finally, the memory stream output C_t serves as the encoder's output, encapsulating the full spectrum of the input time series data, which are then fed into the decoder for further processing. The decoder is assembled using LSTM cells, integrating the memory flow data h_t as the input while initializing with blank input data x_t, facilitating the prediction of the Nino3.4 SST anomaly index at subsequent time point $t+1$. Next, the foreseen Nino3.4 SST anomaly index at $t+1$ is employed as the fresh input data x_{t+1} for the succeeding LSTM timeframe, $t+2$, fostering a continual process to meticulously generate a multistep forecast of the Nino3.4 SST anomaly index spanning the forthcoming two years. This operation is fundamental to the ENSO forecast model grounded in the seq2seq framework. To enhance this structure, a temporal attention mechanism is introduced to delve deeper into the time-dimensional information inherent in the time series data.

3.2 Temporal attention

Within the encoder-decoder framework, the act of condensing all relevant details from the input series into the encoder can invariably lead to information loss. Consequently, the decoder has access only to the data in the static representation vector C_t derived from the encoder side, hindering thorough use of all pertinent information initially available. This limitation can foster a sequential increase in prediction errors as the multistep forecast progresses, culminating in suboptimal outcomes.

Moreover, the importance of various data pieces can fluctuate across different time steps during the prediction process. To counteract this fluctuation and enrich the pool of information accessible to the decoder, a temporal attention mechanism has been integrated into the seq2seq framework. This adaptation is envisioned to enhance the decoder's comprehension of the full temporal sequence data, facilitating more accurate predictions.

The temporal attention mechanism can determine the relevance of information at different time steps for a specific prediction within the temporal dimension. Leveraging this mechanism, the decoder is endowed with the ability to source data from the comprehensive input time series, supplementing the input h_t. The methodology for computing the temporal attention encompasses the subsequent steps. Initially, the temporal attention calculation at time step t incorporates the decoder state y_t along with the full array of encoder-side state vectors $h_1, h_2, \ldots, h_{12}$. Next, attention scores are determined for each individual encoder state h_k. This step entails employing the attention method to gage the correlation score between h_k and the decoder state y_t, as shown in Eq. (7):

$$score(y_t, h_k). \tag{7}$$

Subsequently, the softmax function is used to calculate the attention weights.

$$a_k^{(t)} = \frac{\exp(score(y_t, h_k))}{\sum_{i=1}^{m} \exp(score(y_t, h_i))}, k = 1, \cdots, m. \tag{8}$$

The attention output context vector $c^{(t)}$ is derived as follows:

$$c^{(t)} = a_1^{(t)} h_1 + a_2^{(t)} h_2 + \cdots + a_m^{(t)} h_m = \sum_{k=1}^{m} a_k^{(t)} h_k. \tag{9}$$

Temporal attention facilitates the autonomous selection of the most pertinent time steps from the informational pool for predictive tasks, ensuring a consistent provision of comprehensive input sequence data for each subsequent forecast in a multistep projection. Several approaches are available for deriving attention scores, notably Luong attention and Bahdanau attention (Bahdanau et al. 2014; Luong et al. 2015); the former approach employs a bilinear function for score computation, while the latter approach leverages a multilayer perceptron. In this study, the dot-product strategy is harnessed to compute these crucial scores. Subsequently, the synthesized context vector C_t, bearing important information from the entire input sequence, is fused with the encoder LSTM output y_t, followed by channeling these combined data through a fully connected layer on the encoder end, culminating in the final predictive output.

3.3 ENSO prediction module based on multiscale spatial features

The ENSO phenomenon operates as a hallmark spatiotemporal process characterized by pronounced spatial correlations governing its genesis, evolution, and subsequent phases. Despite the absence of a unified theory within marine science to fully elucidate the physical mechanisms underpinning ENSO phenomena, considerable strides have been made in fostering initial understandings through various lenses, including ocean-atmosphere coupling models, atmospheric oscillations, and oceanic conveyance dynamics. The spatial energy transfer paradigm is central to these endeavors. Consequently, the critical role of spatial information in enhancing the precision of ENSO prognostications becomes evident.

The ENSO prediction model, grounded in the seq2seq framework, engenders a sequence model adept at processing ocean observation datasets, similar to the configurations of two-dimensional gridded data encompassing SST and HC parameters sourced from oceanic and terrestrial regions. Given its intrinsic limitation in handling image data, LSTM requires the direct transformation of two-dimensional data into a one-dimensional vector before integration into the model.

To circumvent this bottleneck, a convolution module, predicated on multi-scale spatial features, is introduced for the adept extraction of spatial characteristics inherent in ocean observation datasets (Szegedy et al. 2015). As depicted in Fig. 2, this module leverages three convolution kernels of distinct dimensions to process the mesh data derived from individual time steps within the input sequence. The ensuing spatial attributes captured at variable scales are concatenated to describe a definitive composite spatial feature set. The 1×1 convolutional kernel serves to reduce dimensionality and computational complexity. This enhanced feature set is subsequently integrated into the ENSO seq2seq prediction model, facilitating enriched training processes.

This innovative multiscale convolution module facilitates diverse receptive fields and has the potential to assimilate a richer tapestry of information across differentiated spatial scales, augmenting the spatial delineation of ocean data. This enhancement, in turn, fortifies the predictive prowess of the ENSO forecasting model. The architecture of this module is shown in Fig. 3.

4 Experiment and analysis

4.1 Experimental platform and data

The proposed model was developed using Python version 3.7.10 and executed on a Tesla K80 GPU. The structural framework of the model was constructed using TensorFlow. This research employed three prominent meteorological datasets: climate model intercomparison project (CMIP), simple ocean data assimilation (SODA), and GODAS, all presented in the prevalent netCDF format. The CMIP and SODA datasets served as the foundation for training the model, while the GODAS dataset was reserved for testing purposes.

Despite variations in the data processing techniques applied to them, all three ocean datasets conform to a uniform final data structure. Each data entry encompasses several meteorological and spatiotemporal parameters, namely, SST and ocean HC, structured with the following dimensions: (year, month, latitude, and longitude). In addition, each monthly data entry in the training set is accompanied by a pertinent label of the Nino3.4 SST anomaly index. In this structured dataset, the initial dimension, representing the 'year', marks the start year of the dataset. The CMIP dataset is extensive, covering a span of 3024 years, incorporating 144 years of historical simulation derived from 21 different coupled model configurations under CMIP. In contrast, the SODA and GODAS datasets contain real observational data, accounting for each year distinctly, spanning 103 years for SODA and 38 years for GODA. To foster a harmonious balance between training and testing data and mitigate potential biases, portions of the initial years in the CMIP and SODA datasets were excluded from the training set. Further elucidating the structure, the second dimension signifies the 'month', with a magnitude of 36. This dimension encapsulates data projected for a three-year period starting from the designated start year.
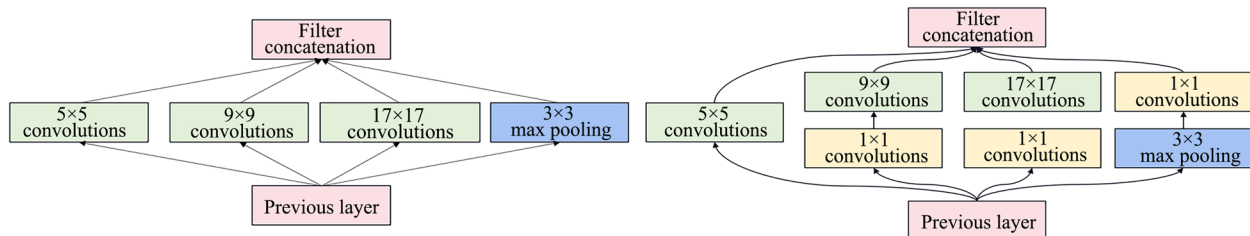
Fig. 2 Multi-scale convolution process: three convolution kernels of different sizes are used to convolve the input and then the resulting spatial features of different scales are sewn together to yield the final hybrid spatial features
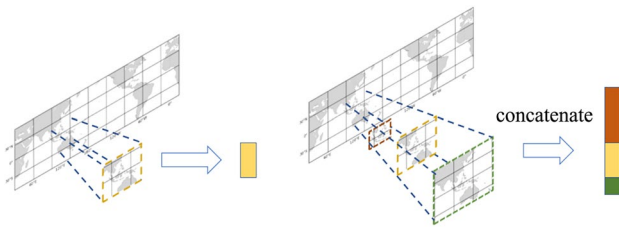
Fig. 3 Multi-scale convolution module: using multiple parallel branches to extract information about features at various scales, this module allows the network to obtain different fields of perception and enrich the spatial features of ocean data

The final two dimensions in each sample delineate the longitude and latitude values, uniformly spanning from $55°$ southern latitude to $60°$ northern latitude and from $0°$ to $360°$ eastern longitude across all datasets. This span equates to a 24×72 grid. Concurrently, the label is aligned with the structure of the initial two dimensions of the training dataset, clarifying the variables associated with the Nino3.4 SST anomaly index and defining them within the (year, month) parameters. The data used in this study diverge substantially from conventional image and sequential data commonly seen in other studies. To facilitate the training process, the data must be transformed into a format amenable to time series analysis, which is characterized by the (samples, time steps, and features) structure. Initial processing entails truncating the second-dimensional segment (month) within each sample, retaining only the data points corresponding to the first 12 months of the relevant year. Subsequently, SST and HC data arrays are harbored within the channel dimension, culminating in the conversion of the terminal three dimensions-latitude, longitude, and channel-into a singular dimensional vector. In parallel, while manipulating the label data, a careful selection of data points within the second dimension (month) is executed, preserving the data spanning the initial 24 months to mirror the predictive output timeline set for a two-year period.

4.2 Experimental design

This study delineates the architecture and hyperparameters of a pioneering ENSO prediction model predicated on a temporal attention mechanism and integrated into a seq2seq framework. This framework leverages two-dimensional SST anomaly and ocean HC datasets, using a sequential span of 12 months encapsulated within the schema (samples, encoder time steps, features). Model optimization is facilitated by employing an Adam optimizer, with the mean square error delineated as the loss function, guided by a learning rate of 0.01 and a batch size of 64. Preliminary training phases are conducted on the CMIP dataset, subsequently advancing to a transference learning stage on the SODA dataset and culminating in evaluative testing on the GODAS dataset.

The encoder component operates with a temporal bandwidth of 12 steps, accommodating encoder tokens configured at dimensions of $24 \times 72 \times 2$. This segment of the model is constructed using a dual-layer LSTM framework, housing 1024 hidden units. Conversely, the decoder module is characterized by a 24-step temporal range and a singular LSTM layer sustaining 1024 hidden units, albeit with a singular decoder token. The hyperparameter configuration for the model is shown in Table 1.

A crucial augmentation in this model is the incorporation of multiscale spatial feature delineation through a convolution kernel module equipped with three diverse kernel dimensions: 5×5, 9×9, and 17×17. This approach necessitates maintaining the original data configuration, which is similar to a two-dimensional graphical representation dictated by (longitude, latitude, channel) parameters. The requisite input data configuration for this experimental framework demands a structure delineated as (samples, time steps, latitude, longitude, channel).

The methodology adopted for training the model leverages the teacher forcing technique, a strategy instituted to mitigate error propagation throughout the sequence generation process. During the training phase, the input furnished to the decoder at each time instance t is predicated on the actual ground truth label corresponding to the preceding time step $t-1$, whereas in the model testing phase, the input received by the decoder at time t corresponds to the (predicted value) output value of the decoder at time $t-1$. In contrast, in the testing phase, the input to the decoder at time t is derived from the model's predicted output from the previous time step $t-1$. To empirically substantiate the efficacy of the proposed model, we employed a widely recognized metric in oceanography, the correlation coefficient. This metric serves as the benchmark for evaluating the predictive prowess of the model, adhering to the computational formula delineated subsequently:

Table 1 Model hyperparameter settings for model experimentation

| Parameter name	Parameter value
Encoder time steps	12
Encoder tokens	$24 \times 72 \times 2$
Encoder LSTM layers	2
Decoder time steps	24
Decoder tokens	1
Decoder LSTM layers	1
Latent dim	1024

$$r(X, Y) = \frac{Cov(X, Y)}{\sqrt{VAR[x] \cdot VAR[y]}}. \qquad (10)$$

In this equation, *Cov* represents the covariance between *X* and *Y* and *VAR*[x] and *VAR*[y] represent the variances of *X* and *Y*, respectively. The magnitude of the correlation coefficient delineates a quantitative measure conveying the extent of the linear relationship between the variables under consideration. A high absolute value signifies a substantial degree of correlation between variables *X* and *Y*, indicating a potent linear association. Conversely, a diminutive absolute value indicates a frail correlation and weaker linear interdependence between *X* and *Y*.

4.3 Prediction results and analysis
In the experimental section of this study, the efficacy of the proposed method was assessed in comparison to several prevalent time series prediction strategies, such as LSTM, CNN+LSTM, ConvLSTM, and SA-ConvL-STM (Shi et al. 2015; Lin et al. 2020). The LSTM model operates using a framework of three integrated LSTM modules. The CNN+LSTM structure integrates three convolutional layers linked to a series of three LSTM layers. Next, the ConvLSTM model incorporates convolution operations directly into the LSTM network, facilitating the simultaneous extraction of spatial and temporal features. Finally, SA-ConvLSTM integrates a self-attention mechanism within the foundational ConvLSTM blueprint to capture extensive spatial dependencies more efficiently.

Figure 4 delineates a comparative analysis of the correlation coefficients obtained from six distinct models: LSTM, CNN+LSTM, ConvLSTM, SA-ConvLSTM, seq2seq attention, and multi-scale convolution seq2seq attention. The data illustrate that the contemporary models considerably enhance the precision in long-term forecasting, as substantiated through an analysis of correlation coefficients with the real and predicted values of the Nino3.4 SST anomaly index spanning a subsequent two-year period. A nuanced assessment of Fig. 4 reveals a converging performance trend among all six models in the initial six-month forecast duration, underscoring their proficiency in encapsulating pertinent short-term ENSO phenomenon indicators. Nonetheless, a discernible decline in forecast accuracy manifests as the timeframe extends, except in the proposed model, which consistently exhibits superior and stabilized performance metrics, underlining its adeptness in negotiating the challenges of long-term time series predictions and effectively harnessing relevant data from the input sequences.
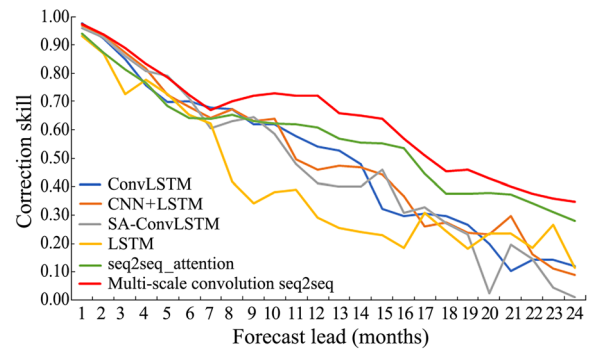
Fig. 4 Nino3.4 index predictions: the correlation coefficients between the predicted and true values of the Nino3.4 SST anomaly index within the next two years show that the multiscale convolution seq2seq attention prediction performs best

Moreover, the introduction of multiscale spatial attributes considerably bolsters the model's six-month predictive accuracy, transcending the results of its counterparts. This enhancement not only refines short-term forecasts but also notably sustains a higher caliber of long-term predictive prowess. In conclusion, the integration of a multiscale convolution module substantively elevates the holistic performance of the predictive model over a two-year prognostic span.

4.4 Transfer learning experiments
The CMIP dataset exhibits substantial biases in data distribution compared with the SODA and GODAS datasets. Although SODA and GODAS use different assimilation algorithms, both datasets are grounded on the assimilation of actual observational data, rendering their data distributions more congruent. The large volume of data available in the CMIP dataset could be considered an advantage; however, exclusive reliance on this dataset for model training may lead to subpar performance when tested on datasets with disparate distributions. To address this limitation, a smaller SODA dataset was employed for transfer learning to enhance the model's alignment with the characteristics of the test data, aiming to improve predictive accuracy. Figure 5 visually contrasts the experimental outcomes before and after applying transfer learning, depicting notable improvement in model efficacy through this strategy.

The experimental outcomes indicate a heightened efficacy in the model's long-term predictive capacity after incorporating SODA reanalysis data into the training regimen. This enhancement substantiates the hypothesis that leveraging a smaller yet more representative dataset for transfer learning can be a potent strategy for mitigating the discrepancy between training and testing data distributions, thereby reducing the bias issue.
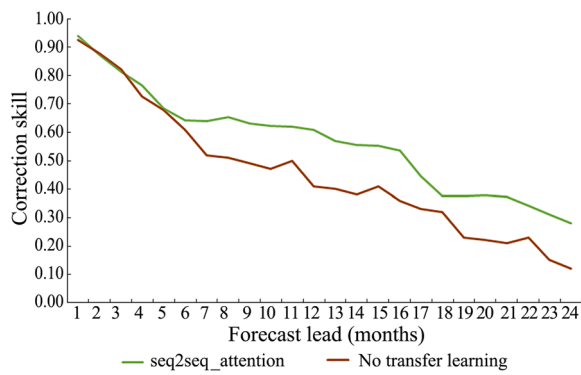
Fig. 5 Transfer learning experiments. The experimental results show that the model's long-term forecasting ability is stronger after migration learning

4.5 Ablation experiments

To substantiate the efficacy of the temporal attention mechanism, a comparative analysis was conducted on the correlation coefficients of the seq2seq model with and without this mechanism. Figure 6 delineates a substantial enhancement in the predictive accuracy for future time series when the temporal attention mechanism is incorporated. This finding affirms that the temporal attention framework adeptly discerns the intricate interrelationships existing between features within a time series. To showcase the potency of multiscale convolution in enhancing the predictive accuracy of the model, an analytical comparison was performed between the seq2seq prediction model using the temporal attention mechanism and an analogous model enriched with a multiscale convolution module. Figure 6 illustrates the outcomes of this empirical examination. A discernible escalation in the correlation coefficient of the predictions delineated in Fig. 6 postulates a marked improvement after integrating
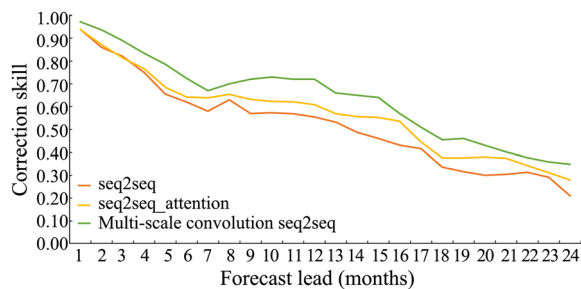
Fig. 6 Ablation experiment. With the addition of the temporal attention mechanism, we see a substantial improvement in the model's ability to predict the future. The addition of the multi-scale convolution module substantially improved the correlation coefficient of the model prediction results

the multiscale convolution module, underscoring its substantial efficacy in optimizing the prediction results. Table 2 shows the average correlation coefficients for the different models. Table 3 shows the average correlation coefficients for different models with different lead times.

5 Conclusions

(1) This study introduces a novel ENSO prediction model grounded on a seq2seq framework fortified with a temporal attention mechanism. The architecture is conceived to analyze time series data harnessed from oceanographic observations, focusing keenly on discerning patterns and tendencies pertinent to ENSO phenomena contained within these data. The overarching ambition is to predict impending trajectories in ENSO alterations over a substantial temporal stretch, thereby enhancing predictive acumen on the SST anomaly index Nino3.4 for subsequent years. The deployment of attention mechanisms in time series forecasting empowers the model to dynamically recognize and prioritize pivotal temporal nodes, thereby fine-tuning the accuracy of ENSO predictions. Through meticulous experimentation, this study corroborates the potency of a seq2seq framework bolstered by temporal attention mechanisms.

(2) Furthermore, this study delineates the advent of a multiscale convolution-centric feature extraction module, marking an innovative stride in the realm of oceanic data processing. This endeavor seeks to decipher the spatial information embedded in ocean meteorological data, fostering a rich repository of spatial characteristics at varying scales. The nucleus of this module is to mine spatial attributes from oceanic observational datasets by deploying convolution kernels of divergent magnitudes, facilitating the extraction of spatial features across diverse scales and enhancing the spatial intel on the features in play. Subsequently, this study unveils an ENSO forecast model leveraging multiscale spatial characteristics and temporal attention mechanisms,

Table 2 Model performance evaluation. Using the average correlation coefficients as the indicator, the performance of the different strategies is shown in table

	seq2seq	+Multi-scale conv	+Attention	+Attention +Multi-scale conv
Cor	0.52	0.55	0.58	0.63

Tao *et al. Intelligent Marine Technology and Systems*   (2024) 2:7

Page 10 of 11

Table 3 Model performance evaluation. Using the average correlation coefficient as an indicator, the performance of the different strategies under different lead times is shown in the table

Forecast lead (months)	1	3	5	7	9	11
Baseline	0.93	0.80	0.65	0.58	0.57	0.57
+Multi-scale conv	0.93	0.81	0.68	0.60	0.61	0.60
+Attention	0.94	0.82	0.68	0.64	0.63	0.62
+Attention +Multi-scale conv	0.97	0.89	0.79	0.67	0.72	0.72

granting it the dexterity to meticulously analyze input sequences across spatial and temporal spectrums, thus opening avenues for markedly improving ENSO forecast model output. The empirical framework designed in this research distinctly showcases the feasibility and ensuing performance enhancement when using spatial information intrinsic to oceanic data through a multiscale convolution-based ENSO predictive module.

Additional information
Edited by: Wenwen Chen

Authors' contributions
Shengen Tao: Conceptualization, Methodology, Formal Analysis, Writing-Original Draft; Yanqiu Li: Software, Validation, Data Curation, Writing-Original Draft; Feng Gao: Methodology, Investigation; Hao Fan: Methodology, Investigation, Funding Acquisition; Junyu Dong: Conceptualization, Resources, Supervision, Funding Acquisition; Yanhai Gan: Methodology, Investigation, Supervision, Data Curation.

Availability of data and materials
Associated data is available at https://github.com/tse1998/ENSO-prediction.

Declarations

Ethics approval and consent to participate
No ethical approval was necessary for this work. Submission declaration and verification: We confirm that our work is original. Our manuscript has not been published, nor is it currently under consideration for publication elsewhere. Shengen Tao, Yanqiu Li, Feng Gao, Hao Fan, Junyu Dong and Yanhai Gan declare that they consent to participate.

Consent for publication
Shengen Tao, Yanqiu Li, Feng Gao, Hao Fan, Junyu Dong and Yanhai Gan declare their consent for publication.

Competing interests
No conflict of interest exists in the submission of this manuscript, and the manuscript is approved by all authors for publication. All the authors listed have approved the manuscript that is enclosed. The authors declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Junyu Dong is the Editor-in-Chief of the journal, but he was not involved in the journal's review of, or decision related to, this manuscript.

References
Adams RM, Chen CC, McCarl BA, Weiher RF (1999) The economic consequences of ENSO events for agriculture. Clim Res 13(3):165–172
Aguilar-Martinez S, Hsieh WW (2009) Forecasts of tropical pacific sea surface temperatures by neural networks and support vector regression. Int J Oceanogr 2009:1–13
Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. Preprint arXiv:1409.0473
Barnston AG, Tippett MK, L'Heureux ML, Li SH, DeWitt DG (2012) Skill of real-time seasonal ENSO model predictions during 2002–11: is our capability increasing? Bull Amer Meteorol Soc 93(5):631–651
Chen D, Cane MA, Kaplan A, Zebiak SE, Huang DJ (2004) Predictability of El Niño over the past 148 years. Nature 428(6984):733–736
Glantz MH (2001) Currents of change: impacts of El Niño and La Niña on climate and society. Cambridge University Press, London
Ham YG, Kim JH, Luo JJ (2019) Deep learning for multi-year ENSO forecasts. Nature 573(7775):568–572
He DD, Lin PF, Liu HL, Ding L, Jiang JR (2019) DLENSO: a deep learning ENSO forecasting model. In: 16th Pacific Rim International Conference on Artificial Intelligence (PRICAI), Cuvu, Yanuca Island, pp 12–23
Jiang JR, Wang TY, Chi XB, Hao HQ, Wang YZ, Chen YR et al (2016) SC-ESAP: a parallel application platform for earth system model. In: 16th Annual International Conference on Computational Science (ICCS), San Diego, pp 1612–1623
Jin EK, Kinter JL, Wang B, Park CK, Kang IS, Kirtman BP et al (2008) Current status of ENSO prediction skill in coupled ocean-atmosphere models. Clim Dyn 31(6):647–664
Lin ZH, Li MM, Zheng ZB, Cheng YY, Yuan C (2020) Self-attention ConvLSTM for spatiotemporal prediction. In: 34th AAAI Conference on Artificial Intelligence, New York, pp 11531–11538
Luo JJ, Masson S, Behera SK, Yamagata T (2008) Extended ENSO predictions using a fully coupled ocean-atmosphere model. J Clim 21(1):84–93
Luong MT, Pham H, Manning CD (2015) Effective approaches to attention-based neural machine translation. Preprint arXiv:1508.04025
Mu B, Peng C, Yuan SJ, Chen L (2019) ENSO forecasting over multiple time horizons using ConvLSTM network and rolling mechanism. In: 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, pp 1–8
Shi XJ, Chen ZR, Wang H, Yeung DY, Wong WK, Woo WC (2015) Convolutional LSTM network: a machine learning approach for precipitation nowcasting. Preprint arXiv:1506.04214

Szegedy C, Liu W, Jia YQ, Sermanet P, Reed S, Anguelov D et al (2015) Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, pp 1–9

Zhang RH (2015) A hybrid coupled model for the Pacific ocean-atmosphere system. Part I: description and basic performance. Adv Atmos Sci 32(3):301–318

Zheng F, Zhu J (2016) Improved ensemble-mean forecasting of ENSO events by a zero-mean stochastic error model of an intermediate coupled model. Clim Dyn 47(12):3901–3915

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.