ST-LSTM-SA：A new ocean sound velocity fields prediction model based on deep learning

Hanxiao YUAN[1], Yang LIU[1,2*], Qiuhua TANG[1,2*], Jie LI[1,2], Guanxu CHEN[1,2], Wuxu CAI[1]

[1] *First Institute of Oceanography, Ministry of Natural Resources, Qingdao* 266061*, China*

[2] *Key Laboratory of Oceanic Surveying and Mapping, Ministry of Natural Resources, Qingdao* 266061*, China*

ABSTRACT

The scarcity of in-situ ocean observations poses a challenge for real-time information acquisition in the ocean. Among the crucial hydroacoustic environment parameters, ocean sound velocity exhibits significant spatial and temporal variability and it is highly relevant to ocean research. In this study, we propose a new data-driven approach, leveraging deep learning techniques, for the prediction of sound velocity fields (SVFs). Our novel spatiotemporal prediction model, ST-LSTM-SA, combines Spatiotemporal Long Short-Term Memory (ST-LSTM) with a self-attention mechanism to enable accurate and real-time prediction of SVFs. To circumvent the limited amount of observation data, we employ transfer learning by firstly training the model using reanalysis datasets, followed by fine-tuning with the in-situ analysis data to obtain the final prediction model. By utilizing the historical 12-months SVFs as input, our model predicts the SVFs for the subsequent 3-months. We compare the performance of five models: Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), Convolutional LSTM (ConvLSTM), ST-LSTM, and our proposed ST-LSTM-SA model in the test experiment spanning from 2019 to 2022. Our results demonstrate that the ST-LSTM-SA model significantly improves the prediction accuracy and stability of sound velocity in both

*Corresponding author: Yang LIU, Qiuhua TANG
Email: yangliu@fio.org.cn, tangqiuhua@fio.org.cn

temporal and spatial dimensions. The ST-LSTM-SA model not only predicts the ocean sound velocity field (SVF) accurately, but also provides valuable insights for spatiotemporal prediction of other oceanic environmental variables.

Key words: sound velocity filed, spatiotemporal prediction, deep learning, self-attention

https://doi.org/10.1007/s00376-024-3219-6

Article Highlights:

- A prediction model for 3D ocean sound velocity fields was developed based on deep learning.
- Employing transfer learning, the ST-LSTM-SA is initially trained on reanalysis data and further refined on in-situ analysis data.
- ST-LSTM-SA shows promising prediction ability by effectively capturing the spatial and temporal variability of sound speed.

1. Introduction

Sound waves are the main medium of underwater information transmission and have various applications in marine engineering, ocean navigation positioning and underwater communication (Akyildiz et al., 2005; Stojanovic and Preisig, 2009). In order to study these applications, it is essential to obtain accurate marine sound environmental parameters. Among them, sound velocity in seawater is one of the parameters that determines the sound propagation characteristics (Kinsler et al., 2000; Heidemann et al., 2012). Sound velocity in seawater is a function of seawater temperature, salinity and pressure, among which the temperature change has the most significant effect on the sound velocity (Chen and Millero, 1977; Mackenzie, 1981). Therefore, sound velocity varies with the ocean dynamic environment in both the time and spatial domains. Due to the vertical stratification of the ocean environment, which in turn makes the sound velocity exhibit a vertically layered structure (Kinsler et al., 2000). In addition, short-term and long-term physical processes in the ocean, such as waves, internal waves, currents and seasonal changes, can alter the marine environment. The superposition of these different periodic physical processes results in complex temporal and spatial variations in sound velocity (Storto et al., 2020).

In current marine research, real-time sound velocity information is predominantly derived from in-situ measurements of sound velocity profiles (SVPs), which capture the variation of sound velocity from the water surface to the seabed (Liu et al., 2023). However, the ocean SVF offers a more comprehensive description of sound velocity distribution in three-dimensional space, which provides a refined representation of the spatial variations of sound velocity. The construction of real-time SVFs is often challenging due to limited observational methods (Dai et al., 2019; Wang et al., 2020). Traditional offshore measurements provide only sparse point-by-point SVPs, which are costly and inefficient to collect frequently. With the development of multiple technology, methods that rely on raw data to predict and invert the sound velocity have been widely studied in recent decades.

Ocean acoustic tomography, systematically introduced by Munk and Wunsch (1979), plays a vital role in marine research, which has paved the way for the development of various SVP inversion methods, including matched acoustic peak arrivals (Skarsoulis et al., 1996) and matched field inversion methods (Tolstoy et al., 1991; Goncharov et al., 1993). Kalman filtering is an optimization algorithm for state estimation and it has been shown to be impactful in ocean forecasting problems (Candy and Sullivan, 1993; Carrière et al., 2009). Compressive sensing (CS) in acoustics has garnered significant attention as an emerging technology in the past decade (Gerstoft et al., 2018). Unlike conventional SVP inversion methods, the CS inversion method effectively estimates fine-scale SVPs through sparse representation using a limited number of SVPs (Bianco and Gerstoft, 2016; Choo and Seong, 2018). Furthermore, machine learning has emerged as an effective way to tackle challenges in marine science, providing fresh avenues for employing data-driven methodologies to make predictions about marine environment (Park and Kennedy, 1996; Jain and Ali, 2006; Chen et al., 2016; Huang et al., 2021). Specifically, a Convolutional Long Short-Term Memory (ConvLSTM) model based on deep learning has been applied into SVP prediction over a three-dimensional sea area, with an average prediction error of less than 1.7 m s^{-1} (Li and Zhai, 2022).

Over the past four decades, researchers have extensively investigated various methods for ocean sound velocity inversion and prediction. Due to the intricate nature of the ocean environment, accurately predicting the ocean SVFs still poses a significant challenge. Traditionally, approaches for spatiotemporal prediction of marine environmental variables rely on ocean numerical simulations, which suffer from significant computational demands, leading to inefficiency in prediction. In fact, the time series of observed ocean data already contains valuable information regarding the internal dynamics and external drivers of the ocean (Espeholt et al., 2022; Shao et al., 2021). Deep learning models, which have the capability to learn from large datasets, can effectively extract the intrinsic characteristics and physical laws inherent in the data (LeCun et al., 2015). As a highly popular and influential technique, deep learning has been successfully applied in various marine prediction researches (Shao et al., 2021; Xiao et al., 2019; Ham et al., 2019; Andersson et al., 2021).

In this research, we propose a new spatiotemporal prediction model (ST-LSTM-SA) for ocean SVFs from a data-driven perspective. Our model combines deep artificial neural networks, including convolutional operations, recurrent neural networks, and self-attention mechanisms, to effectively capture the spatiotemporal variability of sound velocity and enable end-to-end prediction. The model employs an encoding-forecasting network structure that directly outputs future SVFs based on historical observation sequences. During model training, we employ transfer learning by firstly training the model using reanalysis datasets, followed by fine-tuning with the in-situ analysis data to obtain the final prediction model. In terms of accuracy, our model outperforms ANN, LSTM, ConvLSTM and ST-LSTM models, demonstrating superior performance across multiple evaluation metrics, and exhibits enhanced stability in predicting both temporal and spatial dimensions.

2. Data and Data Preprocessing

2.1. Data

The reanalysis dataset is a continuously integrated dataset created by merging observational data with advanced numerical modeling and assimilation techniques (Cummings and Smedstad, 2013). The reanalysis dataset used in this study is the Simple Ocean Data Assimilation System version 2.24, SODA2.24 (Giese and Ray, 2011). This dataset covers the assimilation period of 1871-2008, with a spatial range spanning 0.25°E to 0.25°W and 75.25°S to 89.25°N. It has a horizontal resolution of 0.5°×0.5° and a monthly temporal resolution. The vertical resolution varies from 10 m in the surface layer to 250 m in the bottom layer, divided into a total of 40 unequally spaced vertical layers, available at https://www2.atmos.umd.edu/~ocean/.

The Array for Real-time Geostrophic Oceanography (Argo) program has significantly enhanced oceanic observations, has yielded over 2.5 million ocean profiles to date (Johnson et al., 2022). Starting with these raw observations, researchers have produced numerous in-situ analysis datasets through the application of statistical analyses, optimal interpolation processes, and quality control techniques. (Zhang et al., 2022; Good et al., 2013; Gaillard et al., 2016). In this study, we utilize the Global Gridded Argo Dataset Based on Gradient-Dependent Optimal Interpolation (GDCSM_Argo) (Zhang et al., 2022), covering so far the time range from January 2004 to September 2022 with a monthly temporal resolution. The dataset encompasses the entire global ocean with a horizontal resolution of 1°×1° and consists of 58 unequally spaced vertical layers. The vertical resolution ranges from 5m in the surface layer to 100m in the bottom layer, available at ftp://data.argo.org.cn/pub/ARGO/GDCSM/.

The SODA2.2.4 dataset utilizes a simpler assimilation method and has limited early ocean observation data. In contrast, the GDCSM_Argo dataset is derived from Argo buoy observations, providing an objective representation of the ocean interior. To effectively leverage both datasets, we conducted vertical interpolation on the SODA2.2.4 dataset using cubic spline interpolation, aligning it with the 58 layers of the GDCSM_Argo dataset. Afterward, the pre-training is conducted using the reanalysis dataset, delineated into training, validation, and test sets covering the time spans of 1871-1980, 1981-1994, and 1995-2008, respectively. Subsequent to this, the model undergoes additional training utilizing the GDCSM_Argo dataset. The training set encompasses the time frame from 2004 to 2018, while the test set covers the period of 2019-2022. During this phase, it is noteworthy that there are no alterations made to the hyperparameters of the model (Ham et al., 2019; Pan and Yang, 2010).

2.2. Data preprocessing

1. Data clipping. The dataset is initially cropped to extract the data within the study area range, which spans from 15°S to 15°N and 150°W to 180°W. The study area is shown in Fig. 1 based on ETOPO1 bathymetric model (Amante and Eakins, 2009). This area is situated in the central region of the Pacific Ocean, known for its dynamic climate change and ocean environment.

2. Calculate sound velocity. Reanalysis and in-situ analysis data commonly include variables such as seawater temperature and salinity, which allow for the calculation of sound velocity using empirical equations. To determine sound velocity at each location, the water depth values in the vertical direction were converted from pressure using the pressure-to-depth conversion method proposed by Saunders (1981). Subsequently, the Del Grosso empirical equation (Del Grosso, 1974) for sound velocity was used to calculate the sound velocity information.

3. Data normalization. It involves linearly transforming input data to ensure they are distributed within a specific range. This process helps balance the weights between different features and enhances both the training effectiveness and generalization capability of the model. In our study, we employed the maximum-minimum normalization operation, which scaled all training data to the range of [0, 1]. The calculation procedure for this normalization is as follows:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

where x denotes the sample data, $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the sample data, x^* represents the normalized sample data.

4. Slide sampling. We performed slide sampling on the normalized SVF data using a window size of 15 and a step size of 1. Each sample consists of a sequence of 15 consecutive monthly SVFs. Since the SVFs represent monthly averaged data, we utilize a 12-months input

series with an annual cycle to predict the SVFs for the subsequent 3-months, thereby achieving seasonal forecasting.

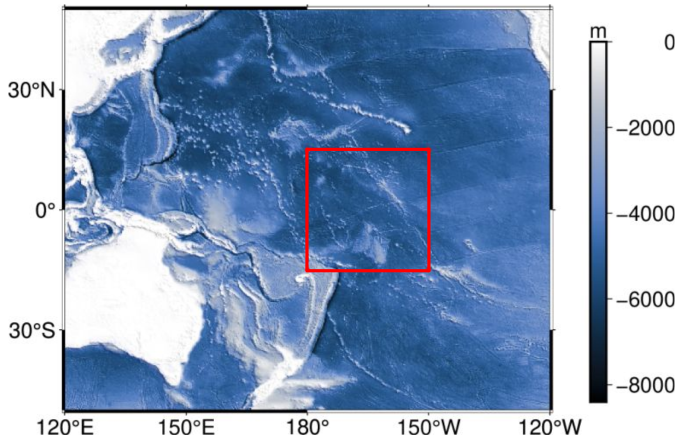


Fig. 1. The bathymetric conditions obtained from ETOPO1 in the central Pacific Ocean, and the study area (150°W-180°W, 15°S-15°N) is delineated by the red box.

3. Methodology

3.1. Problem Definition

In terms of spatial representation, the ocean SVF comprises three dimensions. A three-dimensional grid can be used to represent the spatial location of the sea, wherein each grid cell contains time-dependent sound velocity information. By converting this grid into a tensor, we can express the ocean SVF at a certain time as a three-dimensional tensor $\mathbf{X}_t \in R^{(M \times N \times D)}$, with M, N and D denoting longitude, latitude, and water depth, respectively.

Under the action of complex ocean dynamics processes, the ocean SVFs has obvious time evolution characteristics. Therefore, the prediction problem of SVFs can be regarded as a nonlinear time series prediction problem (Li and Zhai, 2022). Specifically, by leveraging previously observed SVFs series within a given ocean area, we can forecast SVFs for future time intervals with prediction models. Consequently, the temporal prediction problem for the ocean SVFs is defined as follows: constructing a time series $\left(\boldsymbol{X}_{t-n+1}, \boldsymbol{X}_{t-n+2}, \ldots, \boldsymbol{X}_t\right)$ based on n consecutive past SVFs observations to predict the most probable SVFs $\left(\hat{\boldsymbol{X}}_{t+1}, \ldots, \hat{\boldsymbol{X}}_{t+k}\right)$ for the future time range $\left(t+1, \ldots, t+k\right)$ as expressed by the following equation:

$$\hat{\boldsymbol{X}}_{t+1}, \ldots, \hat{\boldsymbol{X}}_{t+k} = f_\theta\left(\boldsymbol{X}_{t+1}, \ldots, \boldsymbol{X}_{t+k} \mid \boldsymbol{X}_{t-n+1}, \boldsymbol{X}_{t-n+2}, \ldots, \boldsymbol{X}_t\right) \quad (2)$$

where f denotes the spatiotemporal prediction model and θ denotes the parameter that is gradually optimized during the training process.

3.2. Basic Deep Learning Models

Long Short-Term Memory (LSTM) is a unique recurrent neural network used for time series problems. It excels at solving long-term dependencies between time series and is applicable to the sound velocity time series prediction problem (Bengio et al., 1994; Hochreiter and Schmidhuber, 1997). However, when it comes to spatiotemporal prediction, the fully connected LSTM networks often struggle to capture spatial features effectively. To overcome this limitation, Shi et al. (2015) introduces the ConvLSTM neural network, which combines convolutional operations with LSTM and has shown success in precipitation nowcasting. The ConvLSTM network incorporates convolutional operations in both input-to-state and state-to-state transitions, allowing for the extraction of spatial features while capturing the dynamic changes of the sequence. In a pioneering study by Li and Zhai (2022), ConvLSTM was applied for the first time to predict SVPs. The experimental results revealed that ConvLSTM outperformed LSTM, providing prediction results that closely aligned with the actual data.

ConvLSTM shares a similar internal structure with LSTM, featuring three gates within each cell: the input gate $\boldsymbol{i}_t$, forgetting gate $\boldsymbol{f}_t$, and output gate $\boldsymbol{o}_t$. The forgetting gate determines the extent to which the previous memory state $\boldsymbol{C}_{t-1}$ is forgotten, while the input gate controls the degree to which the current memory state $\boldsymbol{C}_t$ is updated. The output gate regulates the influence of the current memory state $\boldsymbol{C}_t$ on the output hidden state $\boldsymbol{H}_t$. Figure 2 provides a visual representation of the internal structure of the ConvLSTM cell. The key equations for ConvLSTM are as follows:

$$\boldsymbol{i}_t = \sigma\left(\boldsymbol{W}_{xi} * \boldsymbol{X}_t + \boldsymbol{W}_{hi} * \boldsymbol{H}_{t-1} + \boldsymbol{W}_{ci} \circ \boldsymbol{C}_{t-1} + \boldsymbol{b}_i\right) \tag{3}$$

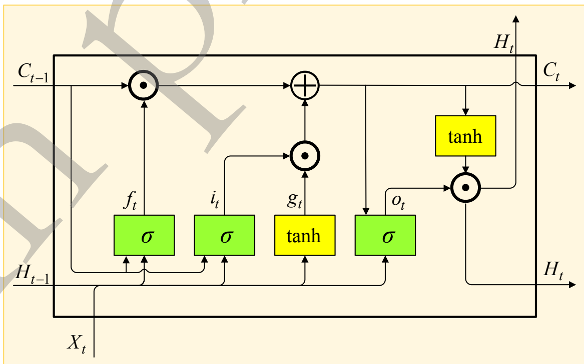
$$\boldsymbol{f}_t = \sigma\left(\boldsymbol{W}_{xf} * \boldsymbol{X}_t + \boldsymbol{W}_{hf} * \boldsymbol{H}_{t-1} + \boldsymbol{W}_{cf} \circ \boldsymbol{C}_{t-1} + \boldsymbol{b}_f\right) \tag{4}$$

$$\boldsymbol{C}_t = \boldsymbol{f}_t \circ \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \circ \tanh\left(\boldsymbol{W}_{xc} * \boldsymbol{X}_t + \boldsymbol{W}_{hc} * \boldsymbol{H}_{t-1} + \boldsymbol{b}_c\right) \tag{5}$$

$$\boldsymbol{o}_t = \sigma\left(\boldsymbol{W}_{xo} * \boldsymbol{X}_t + \boldsymbol{W}_{ho} * \boldsymbol{H}_{t-1} + \boldsymbol{W}_{co} \circ \boldsymbol{C}_t + \boldsymbol{b}_o\right) \tag{6}$$

$$\boldsymbol{H}_t = \boldsymbol{o}_t \circ \tanh\left(\boldsymbol{C}_t\right) \tag{7}$$

Where $\boldsymbol{X}_t$ represents the input at the current time step, $\boldsymbol{W}$ and $\boldsymbol{b}$ are the weight and bias coefficients that are continuously updated during model training. The $*$ symbol denotes the convolution operation, $\circ$ represents the Hadamard (element-wise) operation and σ is the sigmoid activation function.



$\odot$:Element-Wise Product $\oplus$:Element-Wise Addition

Fig. 2. A demonstration of ConvLSTM cell structure.

The ConvLSTM network is a significant advancement in spatiotemporal prediction research, considering the spatial correlation and temporal variation of data. It forms the basis for further studies in this field. An enhanced variant of ConvLSTM, known as the PredRNN network (Wang et al., 2017, 2022), further improves the internal structure to enhance spatiotemporal prediction capabilities. PredRNN introduces a new spatiotemporal LSTM (ST-LSTM) cell, which consists of two memory states: temporal memory state $\boldsymbol{C}_t^l$ and spatiotemporal memory state $\boldsymbol{M}_t^l$. In the ST-LSTM cell, $\boldsymbol{C}_t^l$ is transmitted within adjacent time steps on the same layer, while $\boldsymbol{M}_t^l$ is initially transmitted within the same time step, reaching the top layer at the same moment after passing through the bottom layer at the next moment. This transmission process is illustrated in Fig. 3. This unique method of memory state transfer enables the memory state at the bottom level to depend on both the temporal memory state of the previous moment at same layer and the spatiotemporal memory state from higher layer at historical moments. Consequently, it enhances the interrelation of memory states across different spatial levels. The specific equations of ST-LSTM cells are outlined below:

$$\boldsymbol{g}_t = \tanh\left(\boldsymbol{W}_{xg} * \boldsymbol{X}_t + \boldsymbol{W}_{hg} * \boldsymbol{H}_{t-1}^l + \boldsymbol{b}_g\right) \tag{8}$$

$$\boldsymbol{i}_t = \sigma\left(\boldsymbol{W}_{xi} * \boldsymbol{X}_t + \boldsymbol{W}_{hi} * \boldsymbol{H}_{t-1}^l + \boldsymbol{b}_i\right) \tag{9}$$

$$\boldsymbol{f}_t = \sigma\left(\boldsymbol{W}_{xf} * \boldsymbol{X}_t + \boldsymbol{W}_{hf} * \boldsymbol{H}_{t-1}^l + \boldsymbol{b}_f\right) \tag{10}$$

$$\boldsymbol{C}_t^l = \boldsymbol{f}_t \circ \boldsymbol{C}_{t-1}^l + \boldsymbol{i}_t \circ \boldsymbol{g}_t \tag{11}$$

$$\boldsymbol{g}_t' = \tanh\left(\boldsymbol{W}_{xg}' * \boldsymbol{X}_t + \boldsymbol{W}_{mg} * \boldsymbol{M}_{t-1}^l + \boldsymbol{b}_g'\right) \tag{12}$$

$$\boldsymbol{i}_t' = \sigma\left(\boldsymbol{W}_{xi}' * \boldsymbol{X}_t + \boldsymbol{W}_{mi} * \boldsymbol{M}_{t-1}^l + \boldsymbol{b}_i'\right) \tag{13}$$

$$\boldsymbol{f}_t' = \sigma\left(\boldsymbol{W}_{xf}' * \boldsymbol{X}_t + \boldsymbol{W}_{mf} * \boldsymbol{M}_{t-1}^l + \boldsymbol{b}_f'\right) \tag{14}$$

$$\boldsymbol{M}_t^l = \boldsymbol{f}_t' \circ \boldsymbol{M}_t^{l-1} + \boldsymbol{i}_t' \circ \boldsymbol{g}_t' \tag{15}$$

$$\boldsymbol{o}_t = \sigma\left(\boldsymbol{W}_{xo} * \boldsymbol{X}_t + \boldsymbol{W}_{ho} * \boldsymbol{H}_{t-1}^l + \boldsymbol{W}_{mo} * \boldsymbol{M}_t^l + \boldsymbol{b}_o\right) \tag{16}$$

$$\boldsymbol{H}_t^l = \boldsymbol{o}_t \circ \tanh\left(\boldsymbol{W}_{1\times1} * \left[\boldsymbol{C}_t^l, \boldsymbol{M}_t^l\right]\right) \tag{17}$$

The memory state $\boldsymbol{C}_t^l$ in the ST-LSTM unit follows the gate structures from the ConvLSTM unit, which include the input gate $\boldsymbol{i}_t$ and the forgetting gate $\boldsymbol{f}_t$. Additionally, a new set of input gate $\boldsymbol{i}_t'$ and forgetting gate $\boldsymbol{f}_t'$ are introduced to control the information flow across the memory state $\boldsymbol{M}_t^l$. The output gate $\boldsymbol{o}_t$ is shared by the two memory units to facilitate memory fusion for the storage state $\boldsymbol{H}_t^l$. The input modulation gates $\boldsymbol{g}_t$ and $\boldsymbol{g}_t'$ are involved in the computation of memory states. The coefficients $\boldsymbol{W}$ and $\boldsymbol{b}$ represent the weight and bias terms in the model.

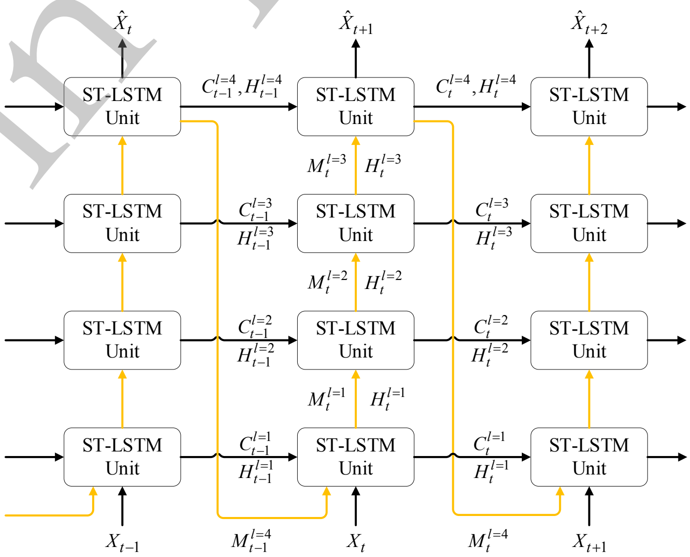


Fig. 3. The memory flow architecture of ST-LSTM, the orange arrows indicate that the spatiotemporal memory state $\boldsymbol{M}_t^l$ is propagated in a zigzag pattern throughout the network, and the black arrows denote the temporal memory state transition paths of $\boldsymbol{C}_t^l$.

3.3. New ST-LSTM-SA Model

Numerous studies have demonstrated that ST-LSTM is well-suited for addressing spatiotemporal prediction problems. Building upon this, we propose a novel ST-LSTM-SA prediction model for SVFs prediction. The architecture of our model follows the encoding-forecasting structure commonly used in earlier studies (Shi et al., 2015, 2017). In our model, we incorporate a self-attention mechanism (Vaswani et al., 2017) between the encoding module and the forecasting module to address temporal dependence issues in the prediction process. This self-attention mechanism dynamically adjusts the weights of the encoding module's outputs at different time steps, enabling us to obtain optimal inputs for the forecasting module at each time step. Figure 4 illustrates the structure of our proposed ST-LSTM-SA model.

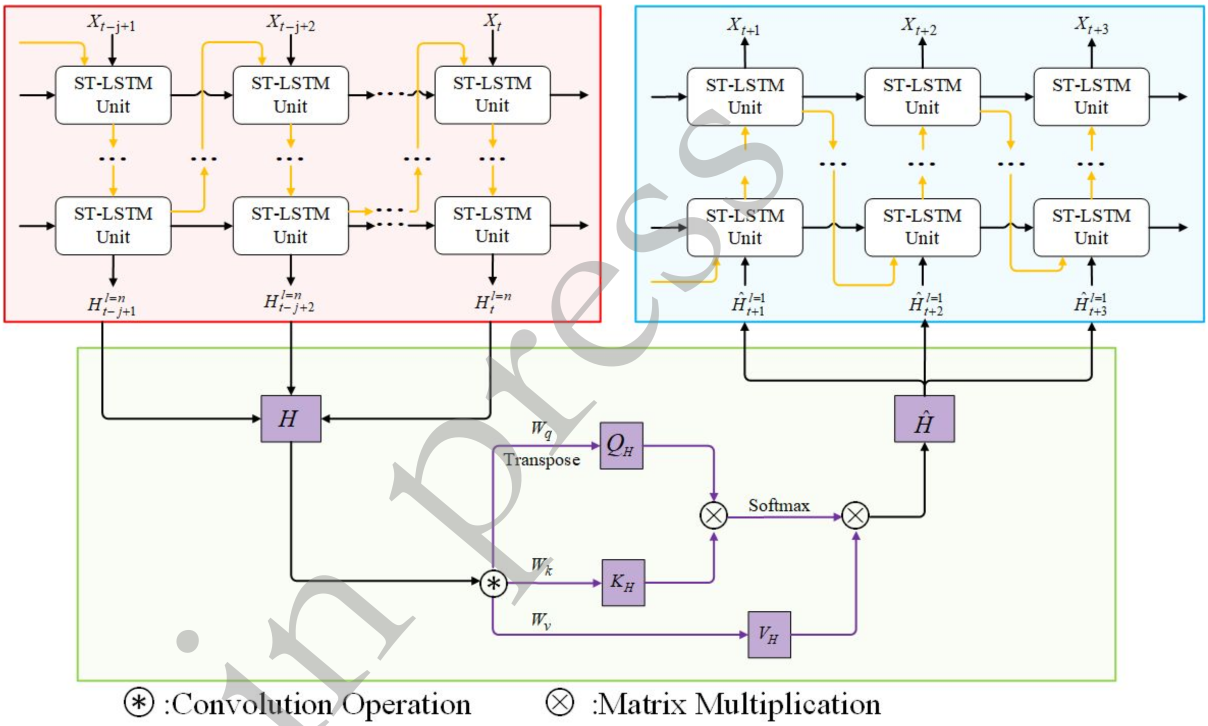


Fig. 4. Structure of the ST-LSTM-SA network. The red block represents the encoding module. The green block presents the attention mechanism module, which demonstrates the operation principle of the self-attention mechanism. The blue block represents the forecasting module.

For the definition of the SVFs prediction problem in Section 3.1, at the current time step t, the model is able to predict the SVFs for the next k time steps based on j historical observations. In the encoding module, highlighted in the red boxed area in the Fig. 4, the input SVF sequence $\left(\boldsymbol{X}_{t-j+1}, \boldsymbol{X}_{t-j+2}, \ldots, \boldsymbol{X}_t\right)$ is encoded by n layers ST-LSTM cells to output j hidden states $\left(\boldsymbol{H}_{t-j+1}, \boldsymbol{H}_{t-j+2}, \ldots, \boldsymbol{H}_t\right)$. The attention mechanism module corresponds to the green boxed area in the Fig. 4, and this part first superimposes the results of the encoding module on the channels to obtain $\boldsymbol{H}$. Then, the query $\boldsymbol{Q}_H$, the key $\boldsymbol{K}_H$ and the value $\boldsymbol{V}_H$ are obtained by mapping to different feature spaces through convolution operations, and $\{\boldsymbol{W}_q, \boldsymbol{W}_k, \boldsymbol{W}_v\}$ represent the weight parameters of the 1×1 convolution operation. The output $\hat{\boldsymbol{H}}$ after attention weight assignment can be obtained by Eq. (18-19):

$$\boldsymbol{\alpha} = \mathrm{softmax}\left(\boldsymbol{Q}_H^T \boldsymbol{K}_H\right) \tag{18}$$

$$\hat{\boldsymbol{H}} = \boldsymbol{\alpha} \boldsymbol{V}_H \tag{19}$$

where $\boldsymbol{Q}_H^T$ denotes the transpose of $\boldsymbol{Q}_H$, softmax is the nonlinear activation function, and $\boldsymbol{\alpha}$ denotes the attention weight distribution located between [0,1]. The forecasting module, depicted within the blue boxed area in the Fig. 4, consists of n layers of ST-LSTM units with the same structure as the encoding module. The output of the attention mechanism module serves as the input for the forecasting module. At the last layer, the forecasting module generates the prediction results for the next k time steps.

3.4. Implementation Details

The experiment was conducted on a server with the following configuration: Windows operating system, 5.10 GHz CPU, 16 GB RAM, and an RTX 3060Ti GPU. The experiment utilized Python 3.8 and the PyTorch 1.11 machine learning framework with CUDA version 11.3 for efficient GPU acceleration. Four neural network models, namely ANN, LSTM, ConvLSTM, and ST-LSTM, were selected and compared with the proposed ST-LSTM-SA model. The purpose of this comparison was to validate the superior performance of the ST-LSTM-SA algorithm.

During the training process, we employed the Adam optimizer (Kingma and Ba, 2014) to optimize the models. Each iteration utilized a batch size of 4, and the initial learning rate was set to 0.001. The learning rate decayed as the number of training sessions increased. To prevent overfitting, regularization was applied to effectively regularize the models. This regularization technique helped enhance the generalization ability of the models and avoid excessive fitting to the training data. Table 1 presents the implementation details for each model. The ANN model was constructed using three fully connected layers and employed the MSE loss function. Due to its limitation in handling only one-dimensional data, the historical SVF sequence needs to be transformed into a one-dimensional tensor for input. The LSTM model followed the encoding-forecasting structure and consisted of four layers of LSTM units with a loss function of MSE. Unlike the ANN model, the LSTM model preserves the time dimension, and the shape of the input tensor is $(B, T, M \times N \times D)$. The ConvLSTM, ST-LSTM, and ST-LSTM-SA models all adopted the encoding-forecasting structure. Each module in these models consisted of two corresponding layers with a uniform hidden state and memory state of 256 channels. The convolutional kernel size was set to 3×3. The loss function employed for these models was MSE loss. The initial input tensor shape for these three models is (B, T, M, N, D). Considering the vertical stratification of ocean sound speed, we sequentially arrange the sound speed values of the nine consecutive bathymetry layers in a 3×3 order, filling insufficient spaces with 0 values. This process results in the input tensor shape of $(B, T, M \times 3, N \times 3, \lfloor D/9 \rfloor)$.

Table 1. Implementation details of models.

Models	Implementation details	Input shape
ANN	Layers=4, hidden_dim=1024	$(B, T \times M \times N \times D)$
LSTM	Encoder: layers=2, hidden_dim=1024 Decoder: layers=2, hidden_dim=1024	$(B, T, M \times N \times D)$
ConvLSTM	Encoder: layers=2, kernel size = (3,3), channels = [256,256] Decoder: layers=2, kernel size = (3,3), channels = [256,56]	$(B, T, M \times 3, N \times 3, \lfloor D/9 \rfloor)$
ST-LSTM	Same as ConvLSTM	$(B, T, M \times 3, N \times 3, \lfloor D/9 \rfloor)$
ST-LSTM-SA	Same as ConvLSTM	$(B, T, M \times 3, N \times 3, \lfloor D/9 \rfloor)$

3.5. Evaluation methods

To evaluate the performance of the different prediction models, we employed several evaluation metrics: root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). RMSE, MAE, and MAPE provide insights into the magnitude of the errors between the predicted and true values, with smaller values indicating better model performance. R^2 measures the strength of correlation between the predicted and true values, with a value ranging from 0 to 1. A value closer to 1 indicates a stronger correlation and better model performance. These metrics are calculated as follows, where $\hat{y}$ and y are the true value and the predicted value, n represents the number of values.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{n}} \tag{20}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|}{n} \tag{21}$$

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{22}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(\bar{y}_i - y_i\right)^2} \tag{23}$$

4. Results and Discussion

4.1. Overall accuracy evaluation

To assess the effectiveness of the proposed algorithm for SVFs prediction, we conducted an analysis and evaluation of the experimental results obtained from the new ST-

LSTM-SA model and four other models. Table 2 presents an overview of the prediction results from different models. The spatiotemporal prediction models (ConvLSTM, ST-LSTM and ST-LSTM-SA), which incorporate convolutional operations, consistently outperform the traditional ANN and LSTM models. The improved performance across all evaluation metrics suggests that the effective extraction of spatial features enhances the accuracy of predicting the three-dimensional structure of the ocean SVFs.

Among the three spatiotemporal prediction models examined in this study, the ST-LSTM model slightly outperforms the ConvLSTM model. This outcome indicates that the introduction of spatiotemporal memory units, which facilitate information exchange across different layers, is crucial for achieving favorable performance. Furthermore, the ST-LSTM-SA model demonstrates further improvement compared to the ST-LSTM model. This finding indicates that the attention mechanism module effectively enhances the quality of information transfer between the encoding module and the forecasting module. By assigning weights to the historical SVF sequences in the encoding module, the attention mechanism module contributes to more realistic predictions from the forecasting module.

Table 2. Overall evaluation indicators for the five models prediction results.

Models	RMSE	MAE	MAPE	R^2
ANN	1.784	1.001	0.066	0.993
LSTM	1.806	1.030	0.068	0.993
ConvLSTM	1.507	0.825	0.055	0.995
ST-LSTM	1.454	0.793	0.052	0.995
ST-LSTM-SA	1.315	0.728	0.048	0.996

Table 3 displays the RMSE and MAPE values for each model's prediction results across different time steps. It is evident that, except for the ANN model, the prediction performance of all models deteriorates over time. This decline can be attributed to the accumulation of errors in the recurrent neural networks used by the other models, whereas the multiple time steps prediction of the ANN model is independent. Notably, the ST-LSTM-SA model consistently outperforms the other models in terms of prediction accuracy. It achieves the lowest prediction errors for the next three months, with reduced increases in prediction errors between adjacent months. Figure 5 presents a statistical histogram of the prediction RMSE for sound velocity. The ST-LSTM-SA model exhibits the smallest error statistics, with approximately 80% of the predicted sound velocity values having the RMSE of less than 1m s^{-1}. This indicates the model's strong spatial and temporal prediction capability, which consistently produces stable and accurate predictions.

Table 3. Statistical of sound velocity prediction error for all test samples in future 3 months.

Mdels	RMSE			MAPE		
	1st month	2nd month	3rd month	1st month	2nd month	3rd month
ANN	1.788	1.789	1.775	0.066	0.066	0.066
LSTM	1.609	1.662	1.758	0.066	0.067	0.068
ConvLSTM	1.432	1.514	1.571	0.053	0.055	0.056
ST-LSTM	1.279	1.478	1.588	0.048	0.053	0.056
ST-LSTM-SA	1.211	1.329	1.399	0.045	0.048	0.051

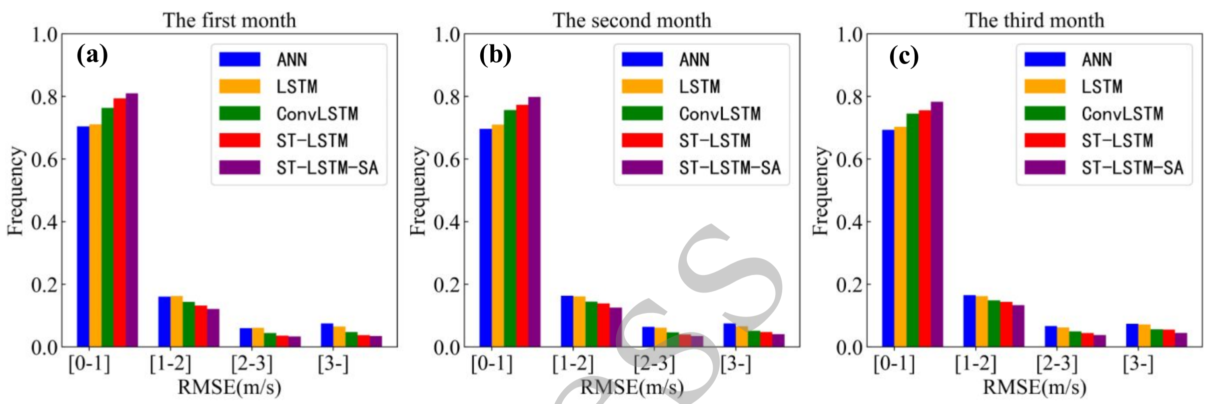


Fig. 5. Frequency distribution of prediction RMSE for all test samples in future 3 months: (a) the first month; (b) the second month; (c) the third month.

In Fig. 6, we present the RMSE of the prediction results at different time steps. For the prediction of the SVFs over the next 31 months, the error curves of the ANN and LSTM models exhibit more pronounced fluctuations. Conversely, the RMSE curves of the three spatiotemporal prediction models demonstrate consistent periodic patterns. Moreover, we observe smoother transitions between adjacent months, leading to a notable reduction in prediction errors. Notably, due to their similar network structures, the error curves of the ST-LSTM and ST-LSTM-SA models closely align with each other. However, in most cases, the ST-LSTM-SA model exhibits further improvement in prediction accuracy compared to ST-LSTM. This finding indicates that the network structure designed in this paper is more suitable for predicting the sound velocity field.

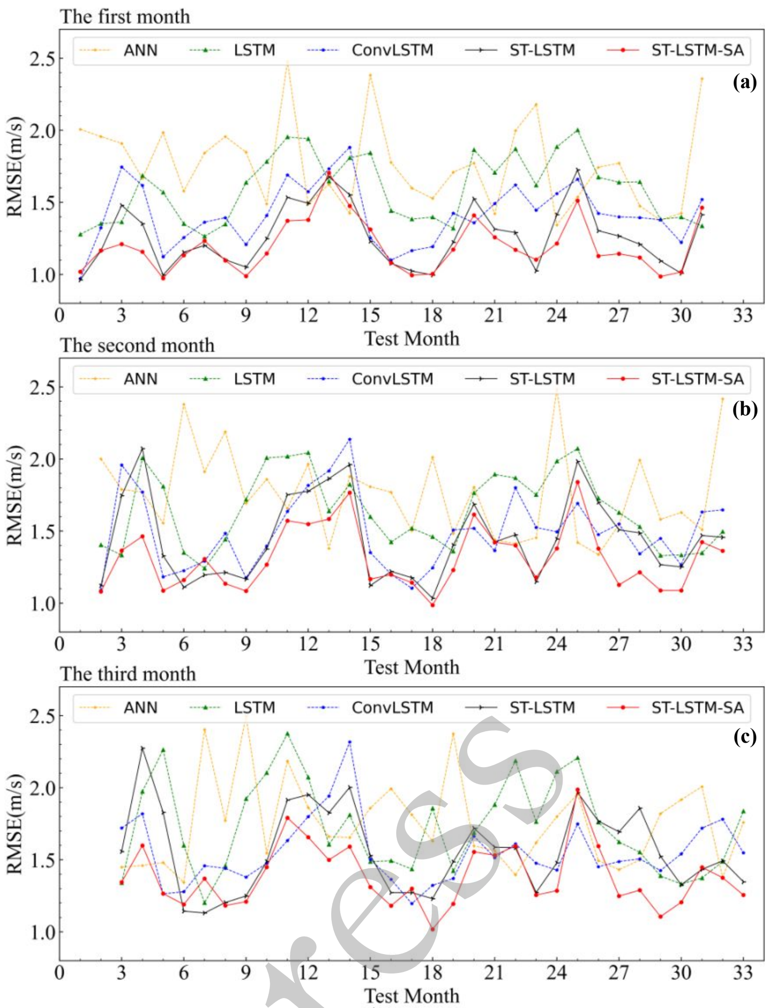


Fig. 6. The RMSE of different models versus the test dataset across different months: (a) the first month; (b) the second month; (c) the third month.

4.2. Spatial predictive accuracy assessment

4.2.1. Horizontal direction analysis

To analyze the predictive accuracy of the models at various spatial locations, we delved into their prediction results along two dimensions: the horizontal direction and the water depth direction. Initially, we computed the RMSE of all water layers at each latitude and longitude grid point for every model. Figure 7 presents a visual representation of the obtained analysis outcomes.

The spatial distribution of errors reveals that the ANN and LSTM models exhibit significant prediction biases across most locations. Interestingly, the spatial distribution of prediction errors remains relatively consistent for the upcoming three months. However, the spatiotemporal prediction models demonstrate noticeable improvements. This improvement can be attributed to the fact that the ANN and LSTM models convert the three-dimensional structure of the SVFs into a one-dimensional representation during training, resulting in a considerable reduction in spatial correlation within the input data. Consequently, these models tend to focus on individual locations rather than the entire SVF during the prediction process.

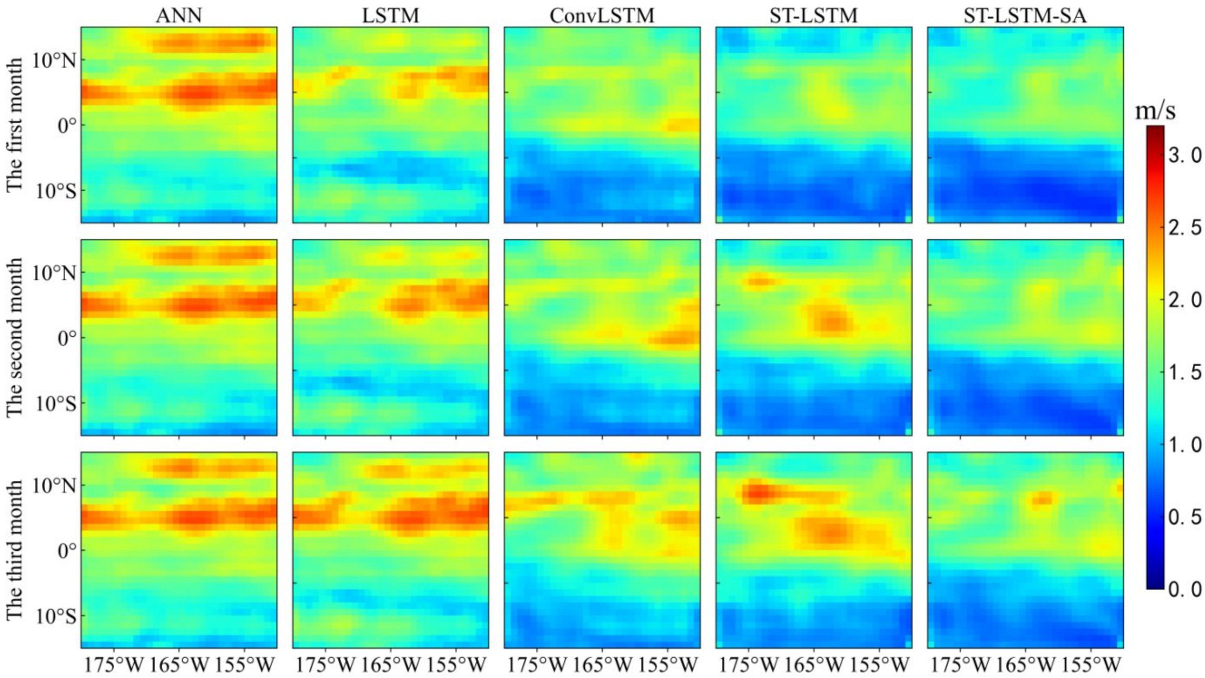


Fig. 7. Statistical predictions in horizontal direction for all test samples in 3 months prediction. Each grid point represents the RMSE of all water layers at one location.

Examining the prediction results of the spatiotemporal prediction models for the subsequent three months, we observe that the prediction error gradually expands over time. All spatiotemporal models exhibit better performance during the first month of the forecast. However, both the ConvLSTM and ST-LSTM models hardly maintain stable prediction capabilities in the following second and third months. In contrast, the prediction results of the ST-LSTM-SA model consistently maintain a balanced spatial distribution, with minimal increases in error between adjacent months. This demonstrates the effectiveness of the ST-LSTM-SA model in capturing both spatial and temporal variations in sound velocity. Moreover, each module within the model plays a distinct role, making it well-suited for predicting the marine sound velocity field.

4.2.2. Depth direction analysis

In order to gain further insights into the models' prediction capabilities at different water depths, we computed the RMSE of the prediction results for each water layer, as illustrated in Fig. 8. It is evident that the prediction errors of all models exhibit a pattern of initially increasing and then decreasing with increasing depth. In other words, the models perform consistently and maintain stable prediction abilities in the surface layer and deep isothermal layer, with the spatiotemporal prediction models achieving a prediction accuracy within 1m s^{-1}. However, in the thermocline layer, the models' prediction accuracy fluctuates significantly. Both the ANN and LSTM models reach maximum RMSE exceeding 4m s^{-1}, while the spatiotemporal prediction model surpasses 3m s^{-1}. This indicates that the ocean environment undergoes more pronounced changes in the thermocline layer. The prediction models struggle to effectively capture the underlying patterns and mechanisms of these changes, as the marine variables such as seawater temperature and salinity are influenced by light and complex physical processes, leading to considerable uncertainties.

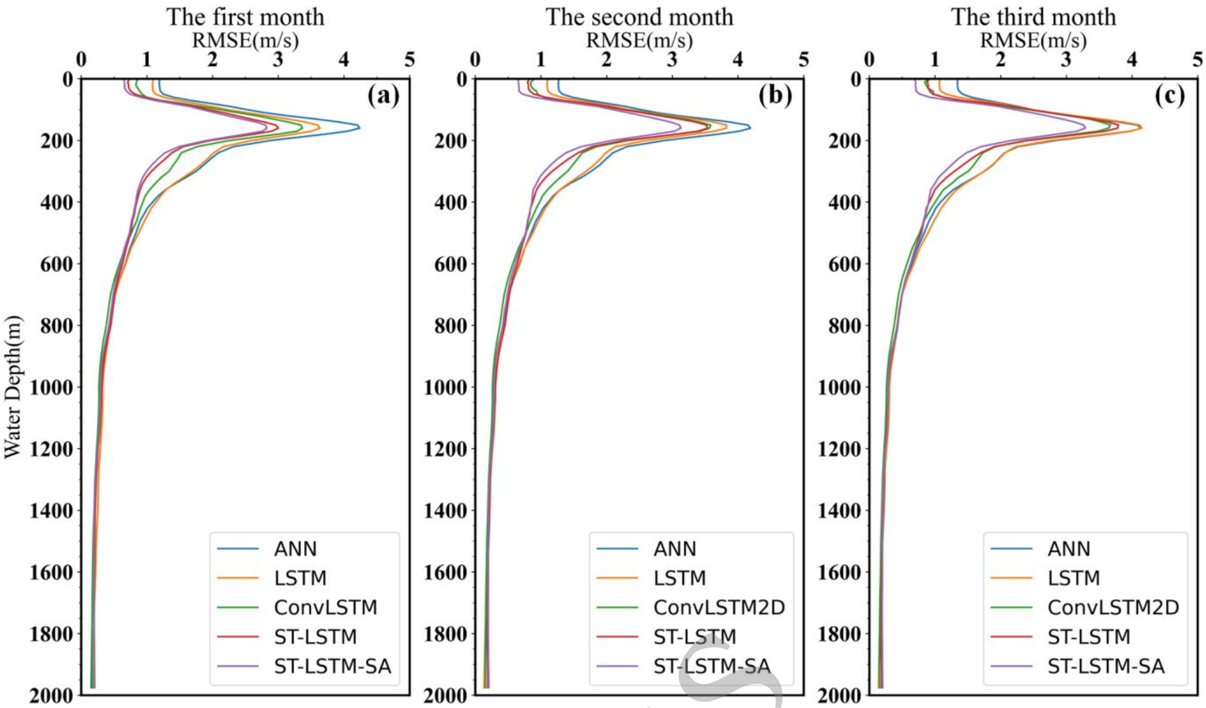


Fig. 8. Statistical predictions on depth direction for all test samples in 3 months prediction: (a) the first month; (b) the second month; (c) the third month.

Although the prediction accuracy of each model in the thermocline layer falls short of expectations, a comparison among the models reveals noteworthy findings. The errors of the spatiotemporal prediction models, unlike those of the ANN and LSTM models, exhibit convergence across different water depths. Notably, the ST-LSTM-SA model demonstrates significant improvement in prediction accuracy for both the surface layer and thermocline layer. This suggests that capturing the spatial characteristics of sound velocity is crucial in addressing the prediction challenges associated with the SVFs. Furthermore, incorporating the attention mechanism enhances not only the prediction accuracy but also the stability of the model across future prediction time.

Figure 9 illustrates the horizontal slices of the sound velocity RMSE for the ConvLSTM, ST-LSTM, and ST-LSTM-SA models at various depths: 50 m, 150 m, 300 m, 500 m, and 800 m. In the surface layer (50 m), the ST-LSTM-SA model demonstrates enhanced prediction accuracy in the central region compared to the other two models. Moving deeper into the water layers at 150 m and 300 m, notable spatial variations in sound velocity prediction errors are observed. To showcase these differences, we select the 11th, 12th, 13th and 14th test months from Fig. 6, where error fluctuations are more pronounced, and present their respective sound velocity horizontal slices at 150 m and 300 m in Fig. 10. From the results, it becomes apparent that sound velocity exhibits significant variability across most of the area north of 5°S at 150 m, while a smoother transition is observed in the southern area. At 300 m, a separation line in sound velocity is still present, but the variation between adjacent months is relatively smoother. Although we acknowledge the potential for doubting the accuracy of the dataset (Zhou et al., 2023), it is important to note that the sharp fluctuations in sound speed primarily impact the model's predictive ability for purely spatiotemporal prediction problems. These fluctuations are closely linked to complex changes in the oceanic environment. As we delve deeper into the water layers at 500 m and 800 m, seawater temperature gradually stabilizes, resulting in regular changes in sound velocity. Consequently, the prediction abilities of the different models become nearly indistinguishable.

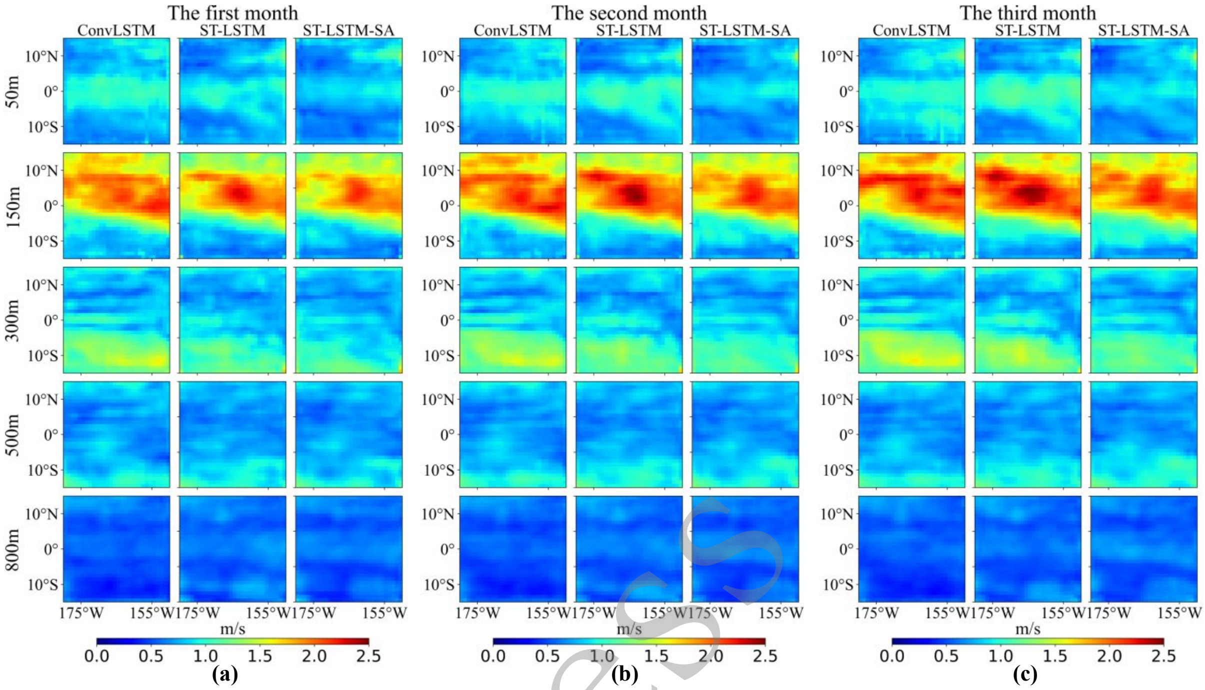


Fig. 9. Horizontal slices of prediction RMSE for all test samples at 50m, 150m, 300m, 500m, and 800m in future 3 months: (a) the first month; (b) the second month; (c) the third month.

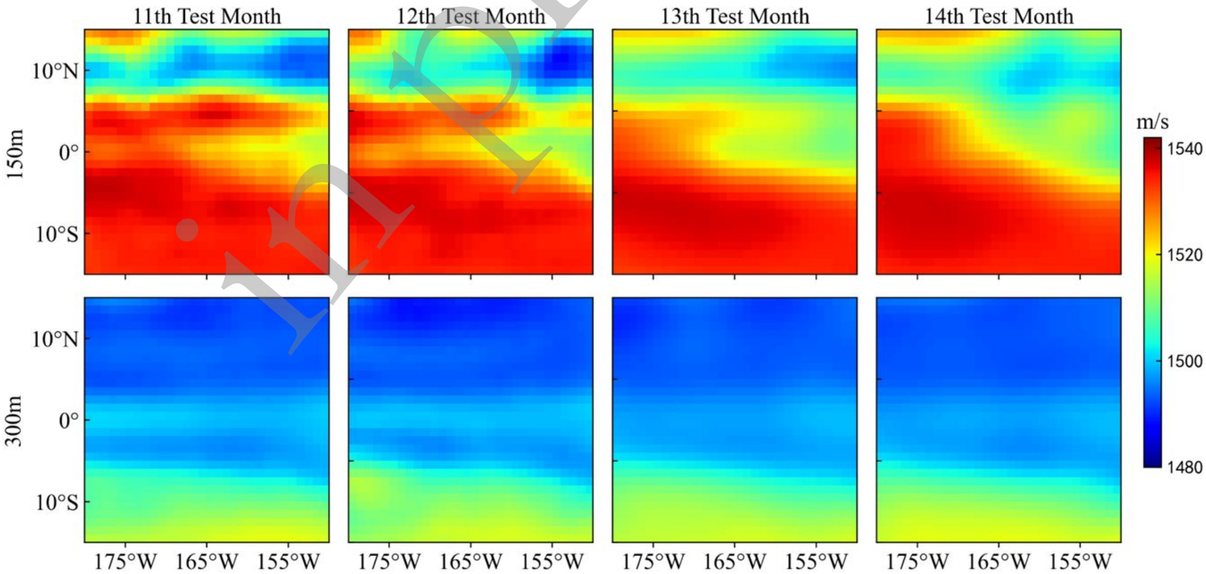


Fig. 10. Horizontal slices of the sound velocity value at 150m and 300m for the 11th, 12th, 13th and 14th test months.

5. Conclusions

Traditionally, numerical ocean simulations are predominantly employed for predicting physical phenomena and internal information within the ocean. This study introduces a novel approach to marine SVFs prediction using the ST-LSTM-SA model, which leverages deep learning techniques. By treating the prediction of SVFs as a nonlinear time series prediction problem and adopting a data-driven approach, this method significantly enhances computational efficiency and reduces resource consumption. The ST-LSTM-SA model is designed to effectively integrate convolutional operations, LSTM, and self-attention mechanisms, allowing for the consideration of both spatial and temporal correlations in the SVF. This enables end-to-end prediction of the SVFs. During model training, transfer learning techniques are employed to train the model weights on different datasets. The SODA2.2.4 reanalysis dataset assists in capturing simple variations in sound velocity over an extended time period, while the GDCSM_Argo in-situ analysis data provides more realistic detailed characteristics of sound velocity, which further refines the model weights.

Through an analysis of the prediction results from January 2019 to September 2022, it is found that the ST-LSTM-SA model outperforms other models across all indicators, demonstrating better agreement with observed results. Temporally, the prediction results of the ST-LSTM-SA model exhibit stability over time, and the self-attention mechanism effectively handles long-term dependencies within the time series. Spatially, traditional ANN and LSTM models convert multi-dimensional data into one-dimensional data during input, disregarding the spatial and temporal correlations in the data, resulting in larger discrepancies in prediction accuracy across different locations. Conversely, the ST-LSTM-SA model demonstrates a more balanced spatial prediction capability, with prediction errors converging across different water depth layers.

Sound velocity in the ocean is influenced by a complex and dynamic environment, making it challenging to accurately describe and simulate its motion and underlying physical laws. In this study, we focus on investigating the spatiotemporal prediction of the sound velocity field. However, there is still room for further improvement in prediction accuracy. We plan to optimize the prediction model by refining its architecture and incorporating new feature data, which is expected to achieve better predictions of the sound velocity field in the future.

Acknowledgments

This work was supported by National Natural Science Foundation of China (42004030), Basic Scientific Fund for National Public Research Institutes of China (2022S03), Science and Technology Innovation Project (LSKJ202205102) Funded by Laoshan Laboratory, and National Key Research and Development Program of China (2020YFB0505805).

Data Availability Statement

We thank National Centers for Environmental Information/National Oceanic and Atmospheric Administration (NCEI/NOAA) for ETOPO1 surface topography data available at https://www.ncei.noaa.gov/access/metadata/landingpage/bin/iso?id=gov.noaa.ngdc.mgg.dem:316, IRI/LDEO Climate Data Library for the SODA version 2.2.4 data available at https://www2.atmos.umd.edu/~ocean/, and Shanghai Ocean University and China Argo Real-time Data Center for the GDCSM_Argo gridded dataset available at https://argo.ucsd.edu/data/argo-data-products/.

References

Akyildiz, I. F., D. Pompili, and T. Melodia, 2005: Underwater acoustic sensor networks: research challenges. Ad hoc networks, **3**(3), 257-279, https://doi.org/10.1016/j.adhoc.2005.01.004

Amante, C., and B. W. Eakins, 2009: ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24[Dataset]. National Geophysical Data Center, NOAA, http://dx.doi.org/10.7289/V5C8276M

Andersson, T. R., J. S. Hosking, M. Pérez-Ortiz, B. Paige, A. Elliott, C. Russell, S. Law, D. C. Jones, J. Wilkinson, T. Phillips, J. Byrne, S. Tietsche, B. B. Sarojini, E. Blanchard-Wrigglesworth, Y. Aksenov, R. Downie, and E. Shuckburgh, 2021: Seasonal Arctic sea ice forecasting with probabilistic deep learning. Nature Communications, **12**(1), 5124, https://doi.org/10.1038/s41467-021-25257-4

Bengio, Y., P. Simard, and P. Frasconi, 1994: Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks, **5**(2), 157-166, https://doi.org/10.1109/72.279181

Bianco, M., and P. Gerstoft, 2016: Compressive acoustic sound speed profile estimation. The Journal of the Acoustical Society of America, **139**(3), EL90-EL94, https://doi.org/10.1121/1.4943784

Candy, J. V., and E. J. Sullivan, 1993: Sound velocity profile estimation: A system theoretic approach. IEEE journal of oceanic engineering, **18**(3), 240-252, https://doi.org/10.1109/JOE.1993.236362

Carrière, O., J. P. Hermand, and J. V. Candy, 2009: Inversion for time-evolving sound-speed field in a shallow ocean by ensemble Kalman filtering. IEEE Journal of Oceanic Engineering, **34**(4), 586-602, https://doi.org/10.1109/JOE.2009.2033954

Chen, C. T., and F. J. Millero, 1977: Speed of sound in seawater at high pressures. The Journal of the Acoustical Society of America, **62**(5), 1129-1135, https://doi.org/10.1121/1.381646

Chen, C., B. Lei, Y. L. Ma, and R. Duan, 2016: Investigating sound speed profile assimilation: An experiment in the Philippine Sea. Ocean Engineering, **124**, 135-140, https://doi.org/10.1016/j.oceaneng.2016.07.062

Choo, Y., and W. Seong, 2018: Compressive sound speed profile inversion using beamforming results. Remote Sensing, **10**(5), 704, https://doi.org/10.3390/rs10050704

Cummings, J. A., and O. M. Smedstad, 2013: Variational data assimilation for the global ocean.

Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II), 303-343, https://doi.org/10.1007/978-3-642-35088-7_13

Dai, M., Y. Li, and K. Yang, 2019: Joint inversion for sound speed field and moving source localization in shallow water. Journal of Marine Science and Engineering, **7**(9), 295, https://doi.org/10.3390/jmse7090295

Del Grosso, V. A., 1974: New equation for the speed of sound in natural waters (with comparisons to other equations). The Journal of the Acoustical Society of America, **56**(4), 1084-1091, https://doi.org/10.1121/1.1903388

Espeholt, L., S. Agrawal, C. Sønderby, M. Kumar, J. Heek, C. Bromberg, C. Gazen, R. Carver, M. Andrychowicz, J. Hickey, A. Bell, and N. Kalchbrenner, 2022: Deep learning for twelve hour precipitation forecasts. Nature communications, **13**(1), 5145, https://doi.org/10.1038/s41467-022-32483-x

Gaillard, F., T. Reynaud, V. Thierry, N. Kolodziejczyk, and K. von Schuckmann, 2016: In situ–based reanalysis of the global ocean temperature and salinity with ISAS: Variability of the heat content and steric height. Journal of Climate, **29**(4), 1305-1323, https://doi.org/10.1175/JCLI-D-15-0028.1

Gerstoft, P., C. F. Mecklenbräuker, W. Seong, and M. Bianco, 2018: Introduction to compressive sensing in acoustics. The Journal of the Acoustical Society of America, **143**(6), 3731-3736, https://doi.org/10.1121/1.5043089

Giese, B. S., and S. Ray, 2011: El Niño variability in simple ocean data assimilation (SODA), 1871–2008. Journal of Geophysical Research: Oceans, **116**(C2), https://doi.org/10.1029/2010JC006695

Goncharov, V. V., and A. G. Voronovich, 1993: An experiment on matched-field acoustic tomography with continuous wave signals in the Norway Sea. The Journal of the Acoustical Society of America, **93**(4), 1873-1881, https://doi.org/10.1121/1.406702

Good, S. A., M. J. Martin, and N. A. Rayner, 2013: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. Journal of Geophysical Research: Oceans, **118**(12), 6704-6716, https://doi.org/10.1002/2013JC009067

Ham, Y. G., J. H. Kim, and J. J. Luo, 2019: Deep learning for multi-year ENSO forecasts. Nature, **573**(7775), 568-572, https://doi.org/10.1038/s41586-019-1559-7

Heidemann, J., M. Stojanovic, and M. Zorzi, 2012: Underwater sensor networks: applications, advances and challenges. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, **370**(1958), 158-175, https://doi.org/10.1098/rsta.2011.0214

Hochreiter, S., and J. Schmidhuber, 1997: Long short-term memory. Neural computation, **9**(8), 1735-1780, https://doi.org/10.1162/neco.1997.9.8.1735

Huang, J., Y. Luo, J. Shi, X. Ma, Q. Q. Li, and Y. Y. Li, 2021: Rapid Modeling of the Sound

Speed Field in the South China Sea Based on a Comprehensive Optimal LM-BP Artificial Neural Network. Journal of Marine Science and Engineering, **9**(5), 488, https://doi.org/10.3390/jmse9050488

Jain, S., and M. M. Ali, 2006: Estimation of sound speed profiles using artificial neural networks. IEEE Geoscience and Remote Sensing Letters, **3**(4), 467-470, https://doi.org/10.1109/LGRS.2006.876221

Kingma, D. P., and J. Ba, 2014: Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, https://doi.org/10.48550/arXiv.1412.6980

Kinsler, L. E., A. R. Frey, A. B. Coppens, and J. V. Sanders, 2000: Fundamentals of acoustics. 3rd ed., John wiley and Sons, 480pp.

LeCun, Y., Y. Bengio, and G. Hinton, 2015: Deep learning. Nature, **521**(7553), 436-444, https://doi.org/10.1038/nature14539

Li, B., and J. Zhai, 2022: A Novel Sound Speed Profile Prediction Method Based on the Convolutional Long-Short Term Memory Network. Journal of Marine Science and Engineering, **10**(5), 572, https://doi.org/10.3390/jmse10050572

Liu, Y., Y. Chen, Z. Meng, and W. Chen, 2023: Performance of single empirical orthogonal function regression method in global sound speed profile inversion and sound field prediction. Applied Ocean Research, **136**, 103598, https://doi.org/10.1016/j.apor.2023.103598

Mackenzie, K. V., 1981: Nine-term equation for sound speed in the oceans. The Journal of the Acoustical Society of America, **70**(3), 807-812, https://doi.org/10.1121/1.386920

Munk, W., and C. Wunsch, 1979: Ocean acoustic tomography: A scheme for large scale monitoring. Deep Sea Research Part A. Oceanographic Research Papers, **26**(2), 123-161, https://doi.org/10.1016/0198-0149(79)90073-6

Pan, S. J., and Q. Yang, 2010: A survey on transfer learning. IEEE Transactions on knowledge and data engineering, **22**(10), 1345-1359, https://doi.org/10.1109/TKDE.2009.191

Park, J. C., and R. M. Kennedy, 1996: Remote sensing of ocean sound speed profiles by a perceptron neural network. IEEE journal of oceanic engineering, **21**(2), 216-224, https://doi.org/10.1109/48.486796

Johnson, G. C., S. Hosoda, S. R. Jayne, P. R. Oke, S. C. Riser, D. Roemmich, T. Suga, V. Thierry, S. E. Wijffels, and J. P. Xu, 2022: Argo—Two decades: Global oceanography, revolutionized. Annual review of marine science, **14**, 379-403, https://doi.org/10.1146/annurev-marine-022521-102008

Saunders, P. M., 1981: Practical conversion of pressure to depth. Journal of Physical Oceanography, **11**(4), 573-574, https://doi.org/10.1175/1520-0485(1981)011%3C0573:PCOPTD%3E2.0.CO;2

Shao, Q., W. Li, G. Han, G. Hou, S. Liu, Y. Gong, and P. Qu, 2021: A deep learning model for forecasting sea surface height anomalies and temperatures in the South China Sea. Journal of Geophysical Research: Oceans, **126**(7), e2021JC017515, https://doi.org/10.1029/2021JC017515

Shi, X., Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems, **28**, https://doi.org/10.48550/arXiv.1506.04214

Shi, X., Z. Gao, L. Lausen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, 2017: Deep learning for precipitation nowcasting: A benchmark and a new model. Advances in neural information processing systems, **30**, https://doi.org/10.48550/arXiv.1706.03458

Skarsoulis, E. K., G. A. Athanassoulis, and U. Send, 1996: Ocean acoustic tomography based on peak arrivals. The Journal of the Acoustical Society of America, **100**(2), 797-813, https://doi.org/10.1121/1.416212

Stojanovic, M., and J. Preisig, 2009: Underwater acoustic communication channels: Propagation models and statistical characterization. IEEE communications magazine, **47**(1), 84-89, https://doi.org/10.1109/MCOM.2009.4752682

Storto, A., S. Falchetti, P. Oddo, Y. M. Jiang, and A. Tesei, 2020: Assessing the impact of different ocean analysis schemes on oceanic and underwater acoustic predictions. Journal of Geophysical Research: Oceans, **125**(7), e2019JC015636, https://doi.org/10.1029/2019JC015636

Tolstoy, A., O. Diachok, and L. N. Frazer, 1991: Acoustic tomography via matched field processing. The Journal of the Acoustical Society of America, **89**(3), 1119-1127, https://doi.org/10.1121/1.400647

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, 2017: Attention is all you need. Advances in neural information processing systems, **30**, https://doi.org/10.48550/arXiv.1706.03762

Wang, J., T. Xu, W. Nie, X. Yu, 2020: The construction of sound speed field based on back propagation neural network in the global ocean. Marine Geodesy, **43**(6), 621-642, https://doi.org/10.1080/01490419.2020.1815912

Wang, Y., M. Long, J. Wang, Z. Gao, and P. S. Yu, 2017: Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. Advances in neural information processing systems, **30**, https://doi.org/10.48550/arXiv.2103.09504

Wang, Y., H. Wu, J. Zhang, Z. Gao, J. Wang, S. Y. Philip, and M. Long, 2022: Predrnn: A recurrent neural network for spatiotemporal predictive learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, **45**(2), 2208-2225, https://doi.org/10.1109/TPAMI.2022.3165153

Xiao, C., N. Chen, C. Hu, K. Wang, J. Gong, and Z. Chen, 2019: Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. Remote Sensing of Environment, **233**, 111358, https://doi.org/10.1016/j.rse.2019.111358

Zhang, C., D. Wang, Z. Liu, S. Lu, C. Sun, Y. Wei, and M. Zhang, 2022: Global Gridded Argo Dataset Based on Gradient-Dependent Optimal Interpolation. Journal of Marine Science and Engineering, **10**(5), 650, https://doi.org/10.3390/jmse10050650

Zhou, G., G. Han, W. Li, X. Wang, X. Wu, L. Cao, and C. Li, 2023: High-resolution gridded temperature and salinity fields from Argo floats based on a spatiotemporal four-dimensional multigrid analysis method. Journal of Geophysical Research: Oceans, e2022JC019386, https://doi.org/10.1029/2022JC019386