Journal of *Marine Science and Engineering*

MDPI

Article

Predictive Modeling of Future Full-Ocean Depth SSPs Utilizing Hierarchical Long Short-Term Memory Neural Networks

Jiajun Lu [1], Hao Zhang [1,2], Pengfei Wu [1], Sijia Li [1] and Wei Huang [1,*]

[1] Faculty of Information Science and Engineering, Ocean University of China, Qingdao 266100, China; lujiajun@stu.ouc.edu.cn (J.L.); zhanghao@ouc.edu.cn (H.Z.); wupengfei@stu.ouc.edu.cn (P.W.); lisijia9430@stu.ouc.edu.cn (S.L.)

[2] Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC V8P 5C2, Canada

* Correspondence: hw@ouc.edu.cn

Abstract: The spatial-temporal distribution of underwater sound speed plays a critical role in determining the propagation mode of underwater acoustic signals. Therefore, rapid estimation and prediction of sound speed distribution are imperative for facilitating underwater positioning, navigation, and timing (PNT) services. While sound speed profile (SSP) inversion methods offer quicker response times compared to direct measurement methods, these methods often focus on constructing spatial sound velocity fields and heavily rely on sonar observation data, thus imposing stringent requirements on data sources. To delve into the temporal distribution pattern of sound speed and achieve SSP prediction without relying on sonar observation data, we introduce the hierarchical long short-term memory (H-LSTM) neural network for SSP prediction. Our method enables the estimation of sound speed distribution without the need for on-site data measurement, significantly enhancing time efficiency. Compared to other state-of-the-art approaches, the H-LSTM model achieves a root mean square error (RMSE) of less than 1 m/s in predicting monthly average sound speed distribution. Its prediction accuracy has improved several-fold over alternative methods, which validates the robust capability of our proposed model in predicting SSP.

Keywords: time series forecast; predictive modeling; full-ocean depth sound speed profile (SSP); hierarchical long short-term memory (H-LSTM)

Citation: Lu, J.; Zhang, H.; Wu, P.; Li, S.; Huang, W. Predictive Modeling of Future Full-Ocean Depth SSPs Utilizing Hierarchical Long Short-Term Memory Neural Networks. *J. Mar. Sci. Eng.* **2024**, *12*, 943. https://doi.org/10.3390/jmse12060943

Academic Editor: Marcello Di Risio

Received: 12 May 2024
Revised: 30 May 2024
Accepted: 2 June 2024
Published: 4 June 2024

1. Introduction

The distribution of underwater sound speed is a critical parameter for underwater positioning, navigation, and timing (PNT) systems, as it directly influences the propagation of underwater acoustic signals [1,2]. With the increasing demand for precision performance in PNT, there is a pressing requirement to quickly and accurately acquire the regional sound speed distributions, and even predict the future sound speed distributions.

The acquisition of an ocean sound speed profile (SSP) primarily involves two methods: SSP measurement and SSP inversion. Measurement can be conducted directly using a sound velocity profiler (SVP) [3], or indirectly through a conductivity, temperature, and depth profiler (CTD) [4,5] or an expendable CTD profiler (XCTD) [6] combined with empirical sound speed equations. However, these methods typically entail lengthy measurement times. For instance, obtaining an SSP at a depth of 2000 m with an SVP or CTD requires at least 80 min [7]. Although using an XCTD can significantly reduce this time to about 20 min for the same depth range, it still lacks real-time capability and is constrained by sensor pressure resistance limits, restricting the measurement depth range.

In recent years, there has been extensive research into ocean SSP inversion methods aimed at rapidly obtaining sound speed distributions. Traditional inversion methods for SSP primarily consist of three categories: matched field processing [8,9], compressed sensing [10–12], and deep learning [13,14]. Most SSP inversion methods focus on the

construction of spatial SSP, and most of them rely on real-time sonar observations and other related data such as temperature and salinity. However, studies focusing on SSP prediction are relatively scarce.

In 1979, Munk and Wunsch introduced the concept of SSP [15,16], laying the groundwork for its inversion by utilizing sound speed propagation time as foundational information. Tolstoy, in 1995, introduced a matching field processing framework for SSP inversion [8], offering an effective solution that circumvents the challenge of establishing a relationship to map sound field distribution to SSPs. In 1995, Markaki et al. introduced a matching field processing method incorporating genetic algorithms (MFP–GA) for SSP inversion [9]. Subsequently, in 1997, Taroudakis integrated matching field processing with modal phase inversion [17]. Yu et al. demonstrated the feasibility of SSP inversion using an MFP–GA method in shallow sea areas in 2010 [18]. Zhang et al. proposed an SSP inversion method based on matched beam processing [19] that was capable of reconstructing SSPs in shallow waters. However, these methods did not achieve satisfactory accuracy in SSP inversion.

To improve the accuracy, Liu et al. introduced an improved method for estimating SSP based on the single empirical orthogonal function (EOF) regression method [20]. Dai et al. introduced an improved particle-filtering technique for sound speed field inversion [21]. Zhang proposed an inversion technique relying on three-dimensional (3D) spatial characteristic sound ray searching and sound propagation time calculation models [22,23]. Zhang et al. presented a method for constructing a time-varying model of regional small-time-scale SSPs based on layered EOFs [24]. In our prior study, we introduced a deep-learning model for SSP inversion [25]. While these methods have improved SSP inversion accuracy, they depend on real-time sonar observation data, thus compromising time efficiency. Recently, Li et al. proposed a self-organizing map (SOM) neural network that incorporates surface sound speed for SSP estimation [26]. This approach achieves SSP construction without relying on sound field observation data, although there remains considerable room for accuracy enhancement. Furthermore, it lacks the ability to predict future sound speed distributions.

While the above various inversion methods can achieve relatively accurate spatial-dimension SSP inversions, they fall short in capturing the temporal evolution of sound speed distribution. Addressing the challenge of SSP prediction in the time domain, we propose a method based on hierarchical long short-term memory (H-LSTM) neural networks. The method mainly learns the change patterns and intrinsic dependencies of the time series of historical sound speed profiles at different sea depths in the spatial domain of the prediction task to realize the precise forecast of future SSP.

The contribution of this research is that we propose an H-LSTM method for the rapid prediction of the future sound speed distribution. The method facilitates accurate predictions of sound speed distributions throughout the entire ocean depth, drawing from historical SSP data. With our SSP prediction method, sound speed distributions can be estimated without the need for on-site data measurements, significantly enhancing time efficiency. We first process the historical sound speed distribution data in layers and establish distinct H-LSTM neural network models for various depth layers. Subsequently, we train and forecast sound speed time series data across different depth layers, ultimately amalgamating the prediction outcomes from each H-LSTM model's layer to construct a full-ocean depth SSP.

The remainder of this paper is structured as follows. Section 2 outlines the H-LSTM prediction method, along with the specific implementation details of the H-LSTM prediction experiment. In Section 3, we present the experimental results and discussions, thoroughly validating the H-LSTM model and comparing its performance with other state-of-the-art methods. Finally, Section 4 provides conclusions drawn from our findings.

2. Methodology

Taking into account seasonal temperature fluctuations, which impact sound speed, and the disparate variations across different depths, we present a novel approach termed H-LSTM for sound speed prediction, leveraging depth stratification. This section begins with an exposition of the conventional LSTM model, followed by an elucidation of our H-LSTM framework.

2.1. Structure of LSTM

2.1.1. LSTM

The long short-term memory (LSTM) neural network is a distinctive variant of recurrent neural networks (RNN) designed to tackle the issues of gradient vanishing and explosion commonly encountered in traditional recurrent architectures. Renowned for its efficacy in time series prediction tasks [27], the LSTM is a formidable solution in this domain.

LSTM regulates information flow through the incorporation of a gating mechanism comprising input gates, forgetting gates, and output gates. Central to LSTM is the "cell state," an internal state capable of reading, writing, and purging information across various time steps in the sequence. The forgetting gate dictates the extent to which information from the preceding cell state C_{t-1} is to be discarded or retained in the current cell state C_t. Meanwhile, the input gate determines the incorporation of current input information into the current cell state C_t. Finally, the output gate governs the transfer of information from the cell state C_t to the estimated value h_t for subsequent time steps. The architectural layout of an LSTM unit is shown in Figure 1.

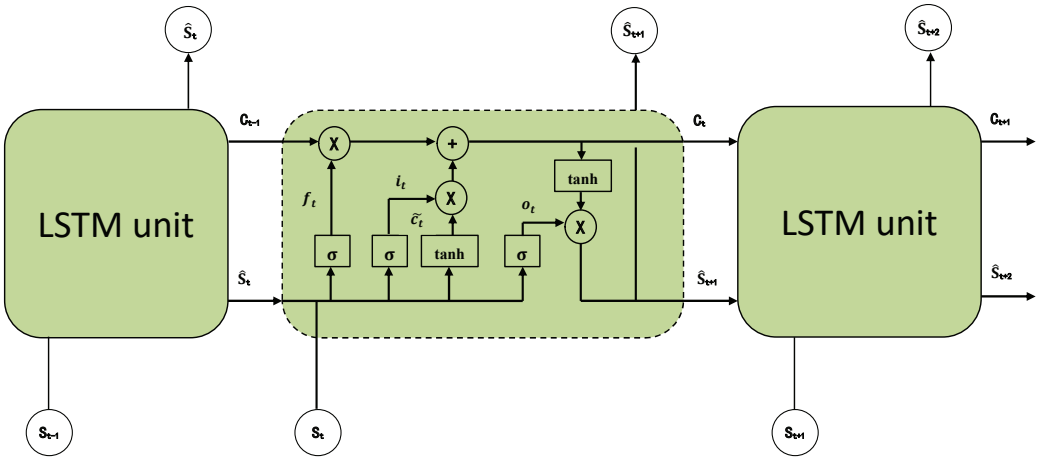


Figure 1. The unit structure of LSTM.

For simplicity, we consolidate multiple LSTM units for clarity, as shown in Figure 2. In the figure, "X" and "+" denote multiplication and addition operations, respectively, while each LSTM unit embodies the same underlying structure as the intermediate unit. σ and *tanh* represent the sigmoid and Tanh activation functions, respectively. Notably, the subscripts t − 1, t, and t + 1 denote the previous, current, and subsequent moments, respectively. Thus, S_{t-1}, S_t, and S_{t+1} denote the SSP input values spanning from moment t − 1 to t + 1, while C_{t-1}, C_t, and C_{t+1} represent the memory units at these respective time instances. Additionally, $\hat{S}_t$, $\hat{S}_{t+1}$, and $\hat{S}_{t+2}$ signify the estimated SSP values of the hidden layer output from moment i − 1 to i + 1. Finally, W and b denote the weight matrices and biases corresponding to the three gate types.

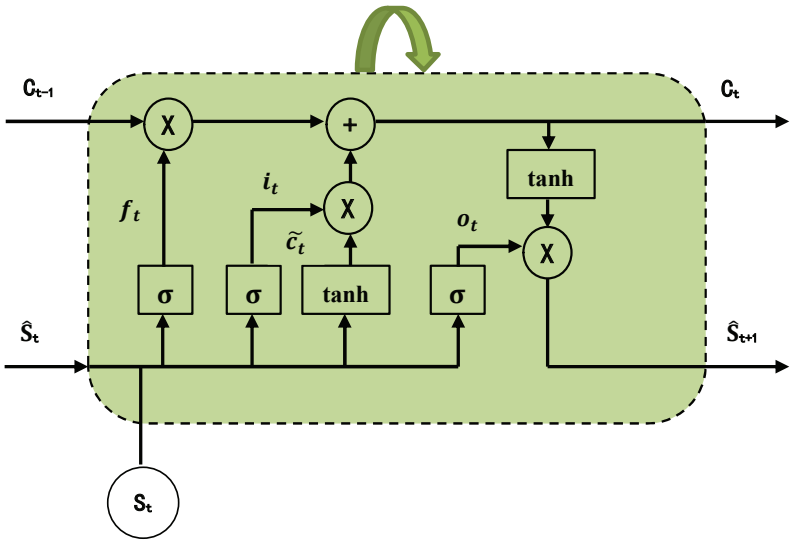


Figure 2. The folding-formed unit structure of LSTM.

The calculation process of the LSTM neural network mainly consists of the following steps.

(1) Calculating output f_t of forgetting gate:

$$f_t = \sigma(W_f \cdot [\hat{S}_t, S_t] + b_f), \tag{1}$$

where W_f represents the weight matrix of the forgetting gate and b_f represents the bias of the forgetting gate;

(2) Calculating output i_t of input gate and candidate cell state $\tilde{C}_t$:

$$i_t = \sigma(W_i \cdot [\hat{S}_t, S_t] + b_i), \tag{2}$$

$$\tilde{C}_t = tanh(W_c \cdot [\hat{S}_t, S_t] + b_c), \tag{3}$$

where W_i and W_c represent the weight matrices of the input gate and candidate cell state inputs, respectively. b_i and b_c represent the biases of the input gate and candidate cell state inputs, respectively;

(3) Updating cell state C_t:

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

(4) Predicting o_t of the output gate and output $\hat{S}_{t+1}$ of the hidden layer:

$$o_t = \sigma(W_o \cdot [\hat{S}_t, S_t] + b_o), \tag{5}$$

$$\hat{S}_{t+1} = o_t \cdot tanh(C_t), \tag{6}$$

where W_o represents the weight matrix of the output gate and b_o represents the bias of the output gate.

2.1.2. H-LSTM

While LSTM excels in time series prediction models, it faces a significant limitation when handling spatiotemporal data. It necessitates the conversion of input features into one-dimensional time series before processing, inevitably resulting in the loss of certain associated feature information during this transformation.

Given the diverse velocity fluctuations across various depths, we introduce the hierarchical long short-term memory neural network model. Our methodology is grounded in the concept of treating sound speed distributions at different depths as distinct, time-varying sequences. To this end, individual H-LSTM models are formulated for each depth layer, trained independently to forecast sound speed values at specific depths for a given

moment. Subsequently, the final full-ocean depth SSP is derived by amalgamating the predicted sound speed values across different depth layers.

In this study, the H-LSTM model comprises an input layer, an H-LSTM layer, a fully connected layer, and an output layer. The architecture of the H-LSTM model is shown in Figure 3, where d_J denotes the Jth depth layer, and H-LSTM$_J$ signifies the H-LSTM model designated for the Jth layer. Before data input, the dataset is stratified based on distinct depth layers and subsequently normalized into a hierarchically standardized SSP dataset. The input hierarchical standardized SSP data undergoes processing within the H-LSTM layer, preserving crucial information while discerning relationships among SSP data across different time points. Introducing a fully connected layer between the H-LSTM layer and the output layer aims to enhance model-fitting performance, culminating in the provision of future SSP prediction outcomes by the output layer.

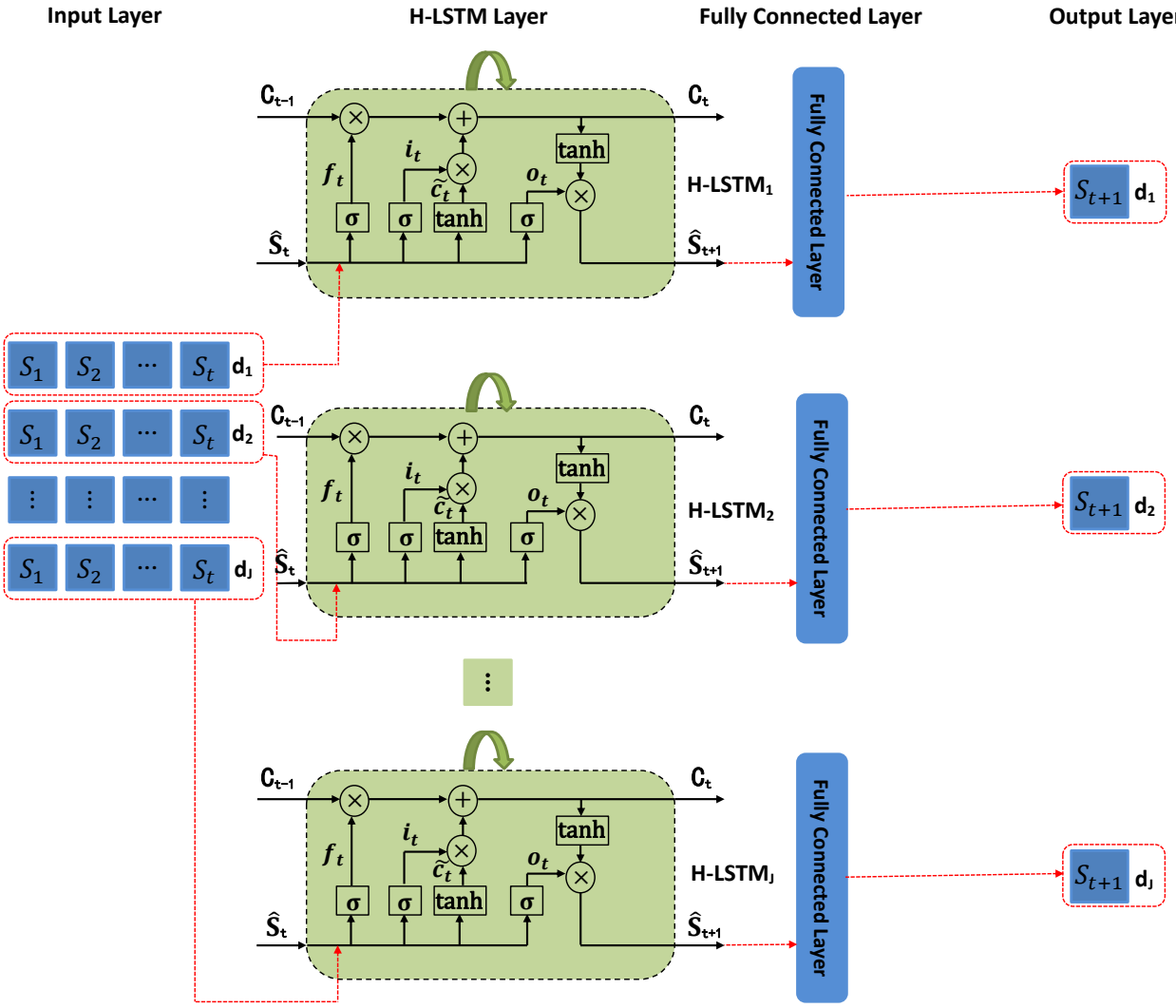


Figure 3. The architecture of the H-LSTM.

2.2. Flowchart of H-LSTM for SSP Prediction

The detailed implementation steps of the sound speed distribution prediction method based on the H-LSTM neural network are shown in Figure 4. These steps encompass dataset preprocessing, H-LSTM neural network building, model training, model validation, and SSP prediction.

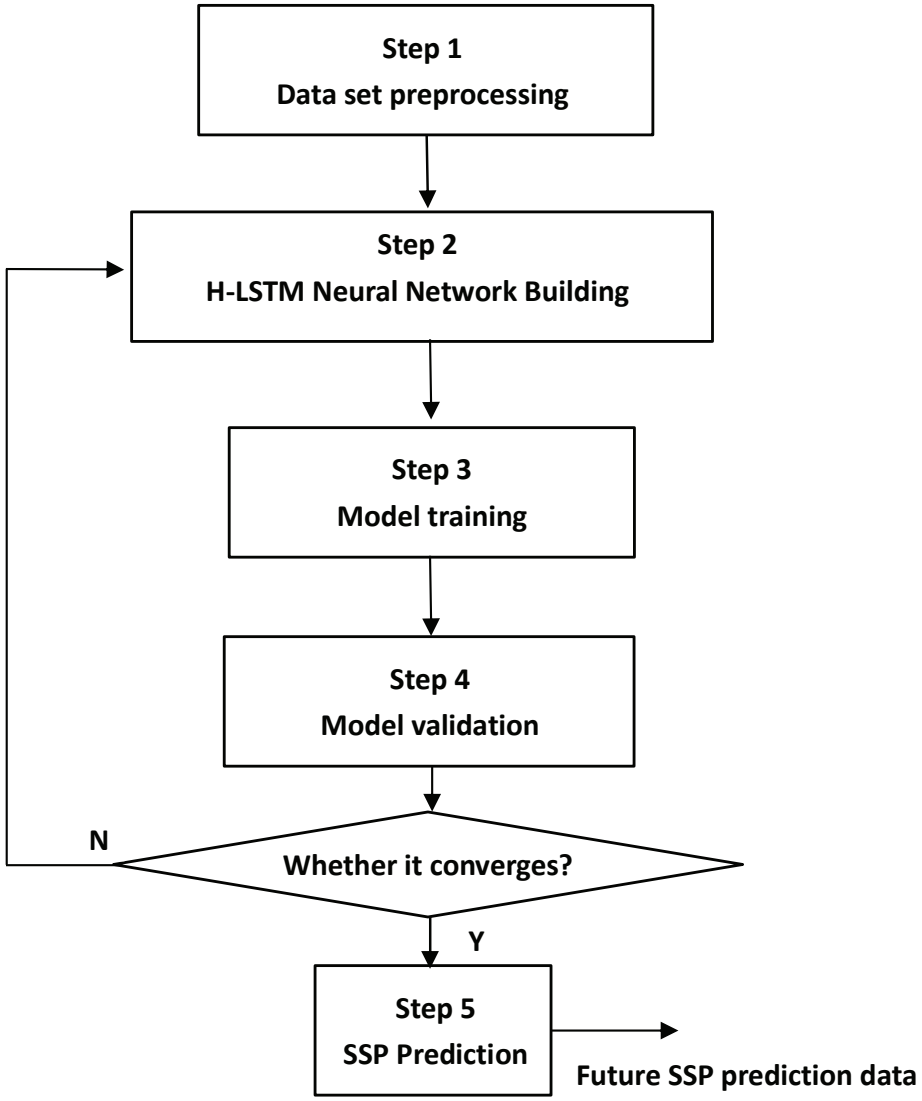


Figure 4. The flow chart of the H-LSTM for SSP prediction.

2.2.1. Dataset Pre-Processing

(1) Data acquisition

Firstly, the research area covers a range of 1 longitude and 1 latitude, according to the specific prediction task of SSP distribution. Historical SSP data within this spatial region are selected from the global Argo dataset as reference profiles. The temporal resolution τ, forecast start time t_0, and maximum forecast duration T_{max} are determined based on the requirements of the sound speed distribution forecasting task.

(2) Data layering by depth

To accommodate the diverse velocity fluctuations at varying depths, the data undergo a layering process before being fed into the H-LSTM model. Initially, the SSP data undergo an unequally spaced layering process along the depth axis, resulting in a total of J layers. The specific unequally spaced layering process is outlined below, taking a depth of 2000 m as an example. From 0 to 10 m, layers are delineated at 5 m intervals; from 10 to 180 m, layers are established at 10 m intervals; from 180 to 460 m, layers are defined at 20 m intervals; from 500 to 1250 m, layers are set at 50 m intervals; from 1300 to 1900 m, layers are assigned at 100 m intervals; depths exceeding 1900 m constitute a single layer. The data are then temporally sorted with a temporal resolution of one month.

The standardized dataset representing the historical hierarchical sound speed distribution is expressed as follows:

$$\mathbf{S} = \mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_i, i = 1, 2, \ldots, I, \tag{7}$$

where $\mathbf{S}_i$ represents the ith SSP. $\mathbf{S}_i$ can be expressed in detail as follows:

$$\mathbf{S}_i = \{s_{i,d_1}, s_{i,d_2}, \ldots, s_{i,d_j}\}^T, j = 0, 1, \ldots, J, \tag{8}$$

where, s_{i,d_j} denotes the sound speed value of the ith SSP at the jth depth layer, with the subscript d_j representing the jth depth layer. The dataset $\mathbf{S}$ in Equation (7) is chronologically ordered, featuring a monthly time resolution. In Equation (8), J serves as the depth layer index label, with d_J representing the maximum common depth layer for the sound speed distribution across the entire sea. Consequently, the dataset $\mathbf{S}$ can be expressed in matrix form using Equations (7) and (8):

$$\mathbf{S} = \begin{bmatrix} s_{1,d_1} & s_{2,d_1} & \cdots & s_{i,d_1} \\ s_{1,d_2} & s_{2,d_2} & \cdots & s_{i,d_2} \\ \cdots & \cdots & \cdots & \cdots \\ s_{1,d_j} & s_{2,d_j} & \cdots & s_{i,d_j} \end{bmatrix}. \tag{9}$$

(3) Data segmentation

To assess and refine the performance of the H-LSTM neural network model, we partitioned the historical standardized dataset representing the hierarchical sound speed distribution over time, denoted as $\mathbf{S}$, into training and validation sets. Taking the minimum duration period of the sound speed data fluctuation as C and predicting the SSP at the time $t+1$ as the target task, the training dataset is the SSP time series data within n cycles from the moment $t+1-nC$ $(t > nC)$ to the moment t, and the validation dataset is the SSP at the moment $t+1$. The hierarchical training set T and the hierarchical validation set V are denoted as follows:

$$\mathbf{T} = \begin{bmatrix} s_{t-nC+1,d_1} & s_{t-nC+2,d_1} & \cdots & s_{t,d_1} \\ s_{t-nC+1,d_2} & s_{t-nC+2,d_2} & \cdots & s_{t,d_2} \\ \cdots & \cdots & \cdots & \cdots \\ s_{t-nC+1,d_j} & s_{t-nC+2,d_j} & \cdots & s_{t,d_j} \end{bmatrix}, \tag{10}$$

$$\mathbf{V} = \begin{bmatrix} s_{t+1,d_1} \\ s_{t+1,d_2} \\ \cdots \\ s_{t+1,d_j} \end{bmatrix}. \tag{11}$$

(3) Data normalization

To expedite convergence and enhance the model's generalization capability, we undertake normalization of the hierarchical training sets. The specific method of data normalization employed is as follows:

$$\tilde{\mathbf{S}} = \frac{S_{i,dJ} - \mu}{\sigma} \tag{12}$$

where $S_{i,dJ}$ represents the time series of SSP at layer J, μ denotes the mean of the corresponding depth layer time series, and σ stands for the standard deviation.

Following normalization, the feature values of the data are constrained to the range $[-1, 1]$, with a mean of 0 and a variance of 1. Normalization is performed by scaling the SSP time series data using the mean and standard deviation, which not only preserves the

distribution shape of the original data but also reduces the impact of outliers on prediction results. Subsequently, we partition the normalized dataset into a training input set and a training output set.

The hierarchical training dataset $\mathbf{T}$ is normalized to $\tilde{\mathbf{T}}$ as follows:

$$\tilde{\mathbf{T}} = \begin{bmatrix} \tilde{s}_{t-nC+1,d_1} & \tilde{s}_{t-nC+2,d_1} & \cdots & \tilde{s}_{t,d_1} \\ \tilde{s}_{t-nC+1,d_2} & \tilde{s}_{t-nC+2,d_2} & \cdots & \tilde{s}_{t,d_2} \\ \cdots & \cdots & \cdots & \cdots \\ \tilde{s}_{t-nC+1,d_j} & \tilde{s}_{t-nC+2,d_j} & \cdots & \tilde{s}_{t,d_j} \end{bmatrix}. \tag{13}$$

Subsequently, the normalized $\tilde{\mathbf{T}}$ is segregated into the hierarchical training input set $\tilde{\mathbf{T}}^{in}$ and the hierarchical training output set $\tilde{\mathbf{T}}^{out}$. We adopt a staggered one-step training approach for the training set, wherein the jth column of data is utilized to predict the $j+1$th column. The hierarchical training input set $\tilde{\mathbf{T}}^{in}$ and the hierarchical training output set $\tilde{\mathbf{T}}^{out}$ are represented as follows:

$$\tilde{\mathbf{T}}^{in} = \begin{bmatrix} \tilde{s}_{t-nC+1,d_1} & \tilde{s}_{t-nC+2,d_1} & \cdots & \tilde{s}_{t-1,d_1} \\ \tilde{s}_{t-nC+1,d_2} & \tilde{s}_{t-nC+2,d_2} & \cdots & \tilde{s}_{t-1,d_2} \\ \cdots & \cdots & \cdots & \cdots \\ \tilde{s}_{t-nC+1,d_j} & \tilde{s}_{t-nC+2,d_j} & \cdots & \tilde{s}_{t-1,d_j} \end{bmatrix}, \tag{14}$$

$$\tilde{\mathbf{T}}^{out} = \begin{bmatrix} \tilde{s}_{t-nC+2,d_1} & \tilde{s}_{t-nC+3,d_1} & \cdots & \tilde{s}_{t,d_1} \\ \tilde{s}_{t-nC+2,d_2} & \tilde{s}_{t-nC+3,d_2} & \cdots & \tilde{s}_{t,d_2} \\ \cdots & \cdots & \cdots & \cdots \\ \tilde{s}_{t-nC+2,d_j} & \tilde{s}_{t-nC+3,d_j} & \cdots & \tilde{s}_{t,d_j} \end{bmatrix}. \tag{15}$$

2.2.2. H-LSTM Neural Network Building

Because SSP prediction is performed layer by layer, we build an H-LSTM network for each depth layer. The model structure of each layer includes the sequence input layer, the H-LSTM layer containing N neurons, and the fully connected layer. The schematic diagram of the model structure has been given in the H-LSTM principle introduction section.

2.2.3. Model Training

During model training, H-LSTM networks at each depth layer undergo separate training processes. We set a different initial dynamic learning rate α_{d_j} and the number of iterations m_{d_j} for each depth layer, where d_j represents the corresponding depth layer. During training of the H-LSTM$_j$ network, the input comprises the time series $\tilde{\mathbf{T}}_{d_j}^{in} = \left[\tilde{s}_{t+1-nC,d_j}, \tilde{s}_{t+2-nC,d_j}, \ldots, \tilde{s}_{t-1,d_j}\right]$, while the model yields the corresponding output time series $\tilde{\mathbf{T}}_{d_j}^{out} = \left[\tilde{s}_{t+2-nC,d_j}, \tilde{s}_{t+3-nC,d_j}, \ldots, \tilde{s}_{t,d_j}\right]$.

2.2.4. Model Validation

(1) Model output

Following model training, the final column of the hierarchical training output set $\tilde{\mathbf{T}}^{out}$ is employed as the model input to forecast future SSP. During layered validation, considering the first depth layer as an illustration, the model input is $\tilde{\mathbf{T}}_{d_1}^{out}(end) = \tilde{s}_{t,d_1}$. The predicted SSP data for the first layer at the subsequent time step is denoted as $P_{d_1} = \hat{s}_{t+1,d_1}$, where $\hat{s}_{t+1,d_1}$ represents the predicted sound speed value in the first depth layer.

By consolidating the predicted results from each layer into a hierarchical column vector, the ultimate predicted hierarchical SSP is represented as $\hat{\mathbf{P}}$:

$$\hat{\mathbf{P}} = \begin{bmatrix} \hat{s}_{t+1,d_1} \\ \hat{s}_{t+1,d_2} \\ \cdots \\ \hat{s}_{t+1,d_j} \end{bmatrix}, j = 1, 2, \ldots, J. \tag{16}$$

(2) Output data denormalization

To facilitate comparison and analysis, the predicted sound speed data undergoes denormalization. This restoration process entails reverting the predicted results $\check{P}$ to the scale range of the original data:

$$\check{\mathbf{P}} = \begin{bmatrix} \check{s}_{t+1,d_1} \\ \check{s}_{t+1,d_2} \\ \cdots \\ \check{s}_{t+1,d_j} \end{bmatrix}, j = 1, 2, \ldots, J, \tag{17}$$

where $\check{s}_{t+1,d_j}$ represents the outcome of reverse-normalizing the predicted sound speed value in the jth depth layer.

(3) Model convergence determination

The root mean square error (RMSE) between the predicted SSP data and the original SSP data serves as the adopted Loss function:

$$RMSE = \sqrt{\frac{\sum(\check{\mathbf{P}} - \mathbf{V})^2}{J}}. \tag{18}$$

If the Loss value tends to be stable, it can be determined that the model has converged and the SSP prediction in step 5 is performed; otherwise, return to step 2.

2.2.5. SSP Prediction

Based on the designated prediction time resolution and the initiation time of the prediction task, we ascertain the number of prediction iterations and identify the historical hierarchical sound speed reference dataset corresponding to the task. Firstly, we use this dataset to train the model and obtain prediction results. Subsequently, we repeat the data denormalization to derive the predicted hierarchical SSP data. Ultimately, we conduct full-ocean depth interpolation on the predicted hierarchical data to obtain future full-ocean depth SSP data.

3. Results and Discussion

3.1. Experiment Settings

3.1.1. Data Source

To validate the efficacy of the H-LSTM SSP prediction method, we selected a specific region highlighted in Figure 5 as the prediction location, situated in the central Pacific Ocean area at coordinates 168.5° E and 16.5° N.

For assessing the performance of our proposed H-LSTM SSP prediction model, we utilized 60 months of measured SSP historical data obtained from the global Argo dataset spanning from 2017 to 2021 at this designated prediction location. The Argo dataset is provided by the China Argo Real-Time Data Center [28]. The Argo dataset comprises a global grid dataset characterized by a spatial resolution of 1° × 1° and a monthly temporal resolution. Details of the data employed in this section are outlined in Table 1.

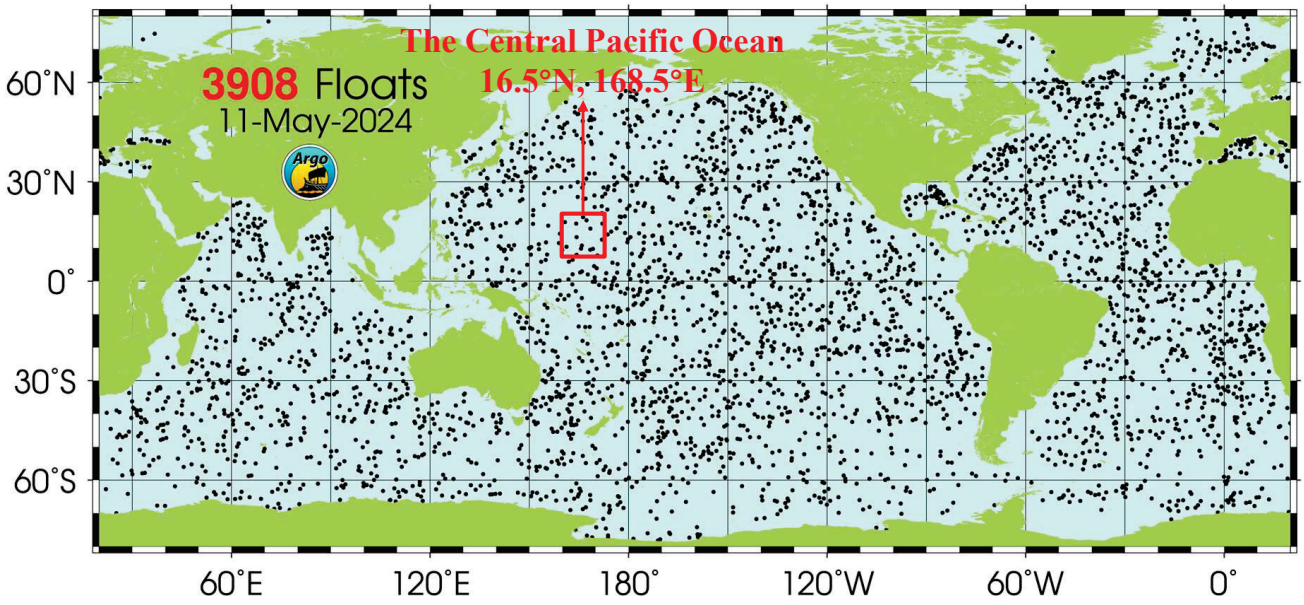


Figure 5. The spatial position of SSP samples.

Table 1. The data information.

Study Area	Input	Time Dimension	Temporal Resolution	Spacial Resolution
168.5° E, 16.5° N	SSP	2017–2021	Month Mean	1° × 1°

3.1.2. Label Data Generation

Before training the H-LSTM prediction model, the data undergo standardization in the depth direction. Following the partitioning method outlined in Section 2, the data are segmented into 58 layers. Subsequently, the data are organized in the time dimension, considering the selected data spanning from 2017 to 2021, encompassing a total of 60 months.

The resulting historical hierarchical sound speed distribution time standardized dataset **S**, after deep stratification and time-sorting processing, can be represented as a sound speed value data matrix with 58 rows and 60 columns:

$$\mathbf{S} = \begin{bmatrix} s_{1,d_1} & s_{2,d_1} & \cdots & s_{60,d_1} \\ s_{1,d_2} & s_{2,d_2} & \cdots & s_{60,d_2} \\ \cdots & \cdots & \cdots & \cdots \\ s_{1,d_{58}} & s_{2,d_{58}} & \cdots & s_{60,d_{58}} \end{bmatrix}. \tag{19}$$

3.1.3. H-LSTM Parameters and Baseline

All data processing procedures conducted in this study were implemented using MATLAB R2023b. In order to ensure the predictive performance of the H-LSTM model, we conducted extensive experiments using a parameter search method, ultimately determining the optimal model parameters for the SSP prediction task. The principal parameter settings of the H-LSTM model utilized in the experiment are summarized in Table 2.

To evaluate the predictive performance of the H-LSTM network model more intuitively, we conducted several experiments comparing it against the mean value prediction method, polynomial fitting method [29], and back propagation (BP) neural network prediction method [30].

Table 2. The principal parameters of the H-LSTM model.

Key Parameters	Settings
Layers	4
Hidden size	128
Initial learning rate	0.01
Epoch	300
Loss function	RMSE

3.2. Effect of Training Dataset on SSP Prediction Performance

To validate the accuracy of the proposed H-LSTM model prediction method in forecasting future SSPs based on learned historical SSP information, we selected 1-year, 2-year, 3-year, and 4-year data segments from the time-standardized dataset **S** as training samples for the model. We assessed the predictive performance of the H-LSTM model under varying learning samples and compared the RMSE between the predicted SSP data and the actual SSP data. The comparison results are shown in Table 3.

Table 3. The effect of training dataset on SSP prediction performance.

Data Set	1 Year	2 Years	3 Years	4 Years
RMSE (m/s)	1.0159	0.5785	**0.4953**	0.4934

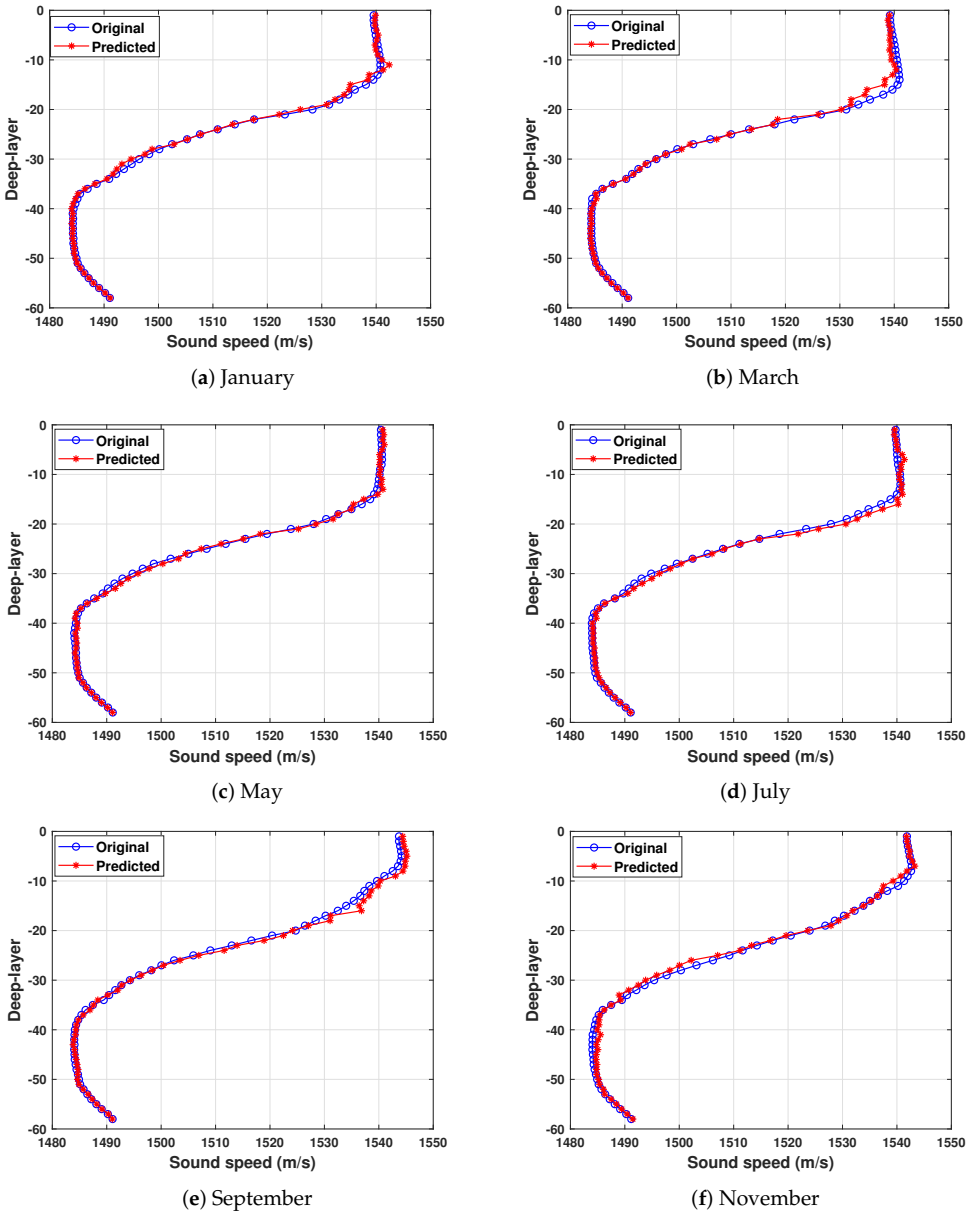
Taking the prediction of January 2021 SSP as an example, Table 3 illustrates the use of historical SSP data from 1–4 years preceding the prediction start as the training set for the H-LSTM model. Analysis of the corresponding RMSE data reveals that, as the size of the training set increases, the model's predictive performance steadily improves. Insufficient training data may hinder the model's ability to capture temporal variation characteristics adequately. Therefore, to ensure accurate prediction performance, it is necessary to train the model utilizing three or more full cycles of historical SSP data prior to the start of the prediction.

3.3. Accuracy Performance of H-LSTM

To validate the accuracy of the model's SSP predictions, sound speed distribution data from the first four complete cycles preceding the prediction start time were extracted from **S**. As an illustration, taking October 2021 as the predicted start time, 48 historical hierarchical sound speed distribution datasets from October 2017 to September 2021 were employed as learning samples for the model. A comparison is shown in Figure 6 between the original and predicted 58-layer sound speed data for selected months in 2021, including January, March, May, July, September, and November.

When comparing the original 58-layer SSP data with the predicted 58-layer SSP data, it becomes clear that the H-LSTM network model can accurately predict future SSP. The predicted sound speed values for each month in 2021 are very similar to the actual sound speed values observed for the corresponding months.

Due to the time-standardized dataset **S**, consisting of data from 58 unequally spaced depth layers rather than the full-ocean deep data, the predicted results cannot be expressed as full-ocean depth SSP. To evaluate the predictive performance of the H-LSTM model for the full-ocean depth SSP, we employed a linear interpolation method to interpolate the hierarchical SSPs as full-depth prediction results across the full ocean depth range, spanning from 0 to 1975 m.



(**a**) January (**b**) March

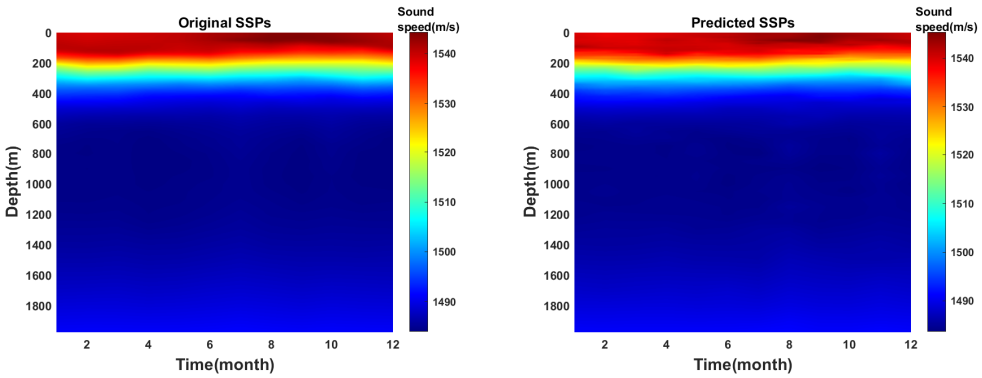
(**c**) May (**d**) July

(**e**) September (**f**) November

Figure 6. The comparison of predicted and original 58-layer SSP for selected months in 2021.

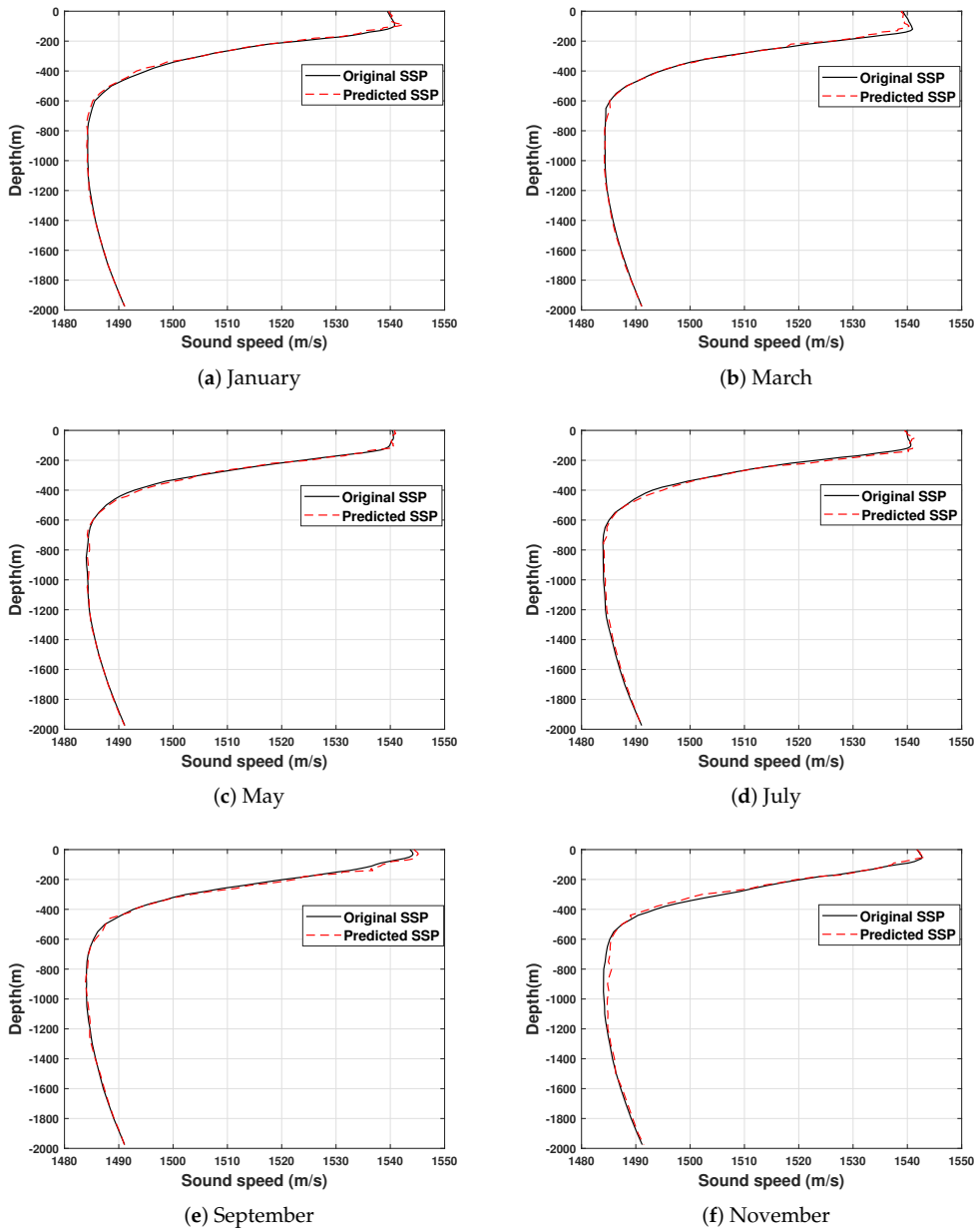
A comparison of the original 12 full-ocean depth SSPs with the predicted 12 SSPs is shown in Figure 7. It can be seen that the 12 monthly averaged full-ocean depth SSPs predicted by the model in the next year are almost perfectly matched with the actual SSPs, and the fitting effect is excellent, accurately reflecting the characteristic information of the actual SSP.

To further visually compare the differences between the predicted and actual SSPs, a two-dimensional comparison of SSPs for selected months is shown in Figure 8, including January, March, May, July, September, and November. It becomes evident that the H-LSTM network model can make precise forecasts of future full-ocean depth SSPs. The predicted future SSPs accurately capture the principal characteristics of the actual SSPs.



(**a**) The original full-ocean depth SSPs (**b**) The predicted full-ocean depth SSPs

Figure 7. The comparison of the original 12 full-ocean depth SSPs with the predicted 12 SSPs.



(**a**) January (**b**) March

(**c**) May (**d**) July

(**e**) September (**f**) November

Figure 8. The comparison of predicted and original full-ocean depth SSP for selected months in 2021.

The Table 4 provides an overview of the prediction errors for each month in 2021, as indicated by the RMSE between the predicted full-ocean depth SSP and the actual full-ocean depth SSP.

Table 4. The prediction errors for each month in 2021.

Predicted Area	**Pacific Ocean**
Depth	0–1975 m
Year	2021
Month	RMSE (m/s)
1	0.4934
2	1.1310
3	0.6630
4	0.7163
5	0.4586
6	0.4618
7	0.9990
8	0.6212
9	0.6955
10	0.5565
11	0.8606
12	0.6443
Average Result	0.6917

3.4. Performance Comparison with Other Methods

To comprehensively validate the feasibility of our proposed H-LSTM neural network model for predicting future SSPs, we conducted several comparative experiments, contrasting it with the mean value prediction method, polynomial fitting method, and BP neural network prediction method.

In these comparative experiments, considering the prediction of October 2021 as an example, the H-LSTM model utilized 58 layered SSP data from October 2017 to September 2021 as a training set to predict future SSPs. The mean value prediction method computed the mean of four consecutive years of historical SSP data preceding October 2021 as the predicted future SSP data. For the polynomial fitting method, polynomial fitting was performed on the historical SSP data of two consecutive years before October 2021 to obtain the predicted future SSP. Utilizing data from two consecutive periods before the prediction start time serves two purposes. Firstly, the polynomial fitting does not necessitate an extensive dataset for SSP prediction and secondly, it distinguishes this method from the mean value prediction method. The BP neural network employed the same training dataset as the H-LSTM model, using it as the model's training set to predict future SSPs.

To compare the prediction performance of the various methods, we present a comparison between the predicted full-ocean depth SSP and the actual SSP. Figure 9 illustrates the comparison of the four methods predicted SSPs.

From Figure 9, it is apparent that: the mean value prediction method yields better SSP prediction results in the deep-sea part (below 900 m), but exhibits relatively poor prediction results above 900 m. The polynomial fitting method can roughly capture the main features of future SSPs smoothly, albeit with some numerical deviations overall. The SSPs predicted by the BP neural network method lack smoothness and repeatedly fluctuate around the real SSP, failing to capture the main features of the actual SSP. This method is not suitable for predicting future SSPs. In contrast, the proposed H-LSTM neural network method maintains the main features of the actual SSP without significant numerical deviations. Both shallow and deep-sea areas show relatively good prediction results, with only minor deviations in certain details.

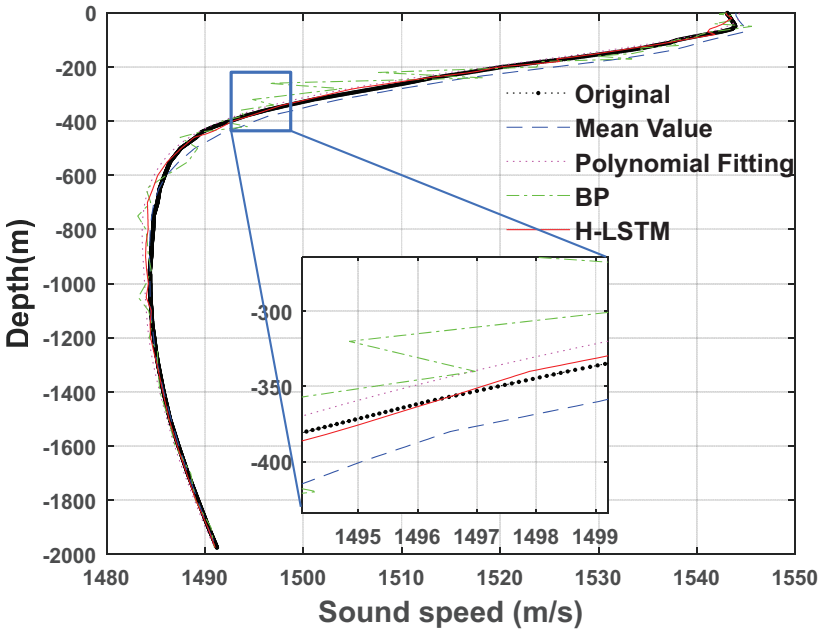


Figure 9. The comparison between predicted and original full-ocean depth SSP.

To compare the performance of several methods more clearly in predicting future SSPs at full-ocean depths, we selected a four-layer structure of data from different depth regions: the surface layer, the seasonal thermocline, the main thermocline, and the deep-sea isothermal layer of marine SSP. The depths corresponding to the error data include 5 m, 10 m, 20 m, 50 m, 70 m, 100 m, 200 m, 500 m, 800 m, 1200 m, 1500 m, and 1800 m. These depths encapsulate a typical four-layer structure of SSP and broadly cover full-ocean depth SSP data. Figure 10 illustrates the comparison of prediction errors of four methods for predicting future SSPs at full-ocean depths.

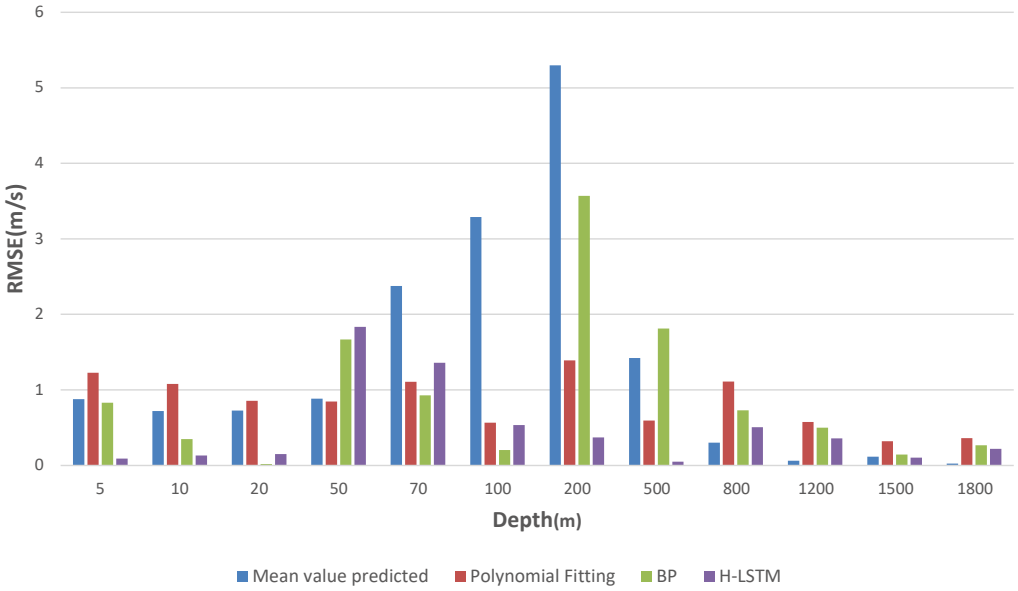


Figure 10. The comparison of prediction effects at different depths.

In the mean value prediction method, the maximum prediction error occurs at a depth of 200 m, exceeding 5 m/s. For the polynomial fitting, the prediction error is relatively stable at each depth layer, with a maximum error of no more than 1.5 m/s. However, in most depth layers, errors surpass 1 m/s. Similarly, the BP neural network prediction method

exhibits a maximum error at a depth of 200 m, exceeding 3 m/s. Conversely, the proposed H-LSTM neural network prediction method demonstrates significantly smaller prediction errors. The maximum error does not exceed 2 m/s, with only two layers experiencing errors surpassing 1 m/s, and the minimum error falling below 0.5 m/s. In summary, H-LSTM outperforms other methods in predicting future SSPs at full-ocean depth.

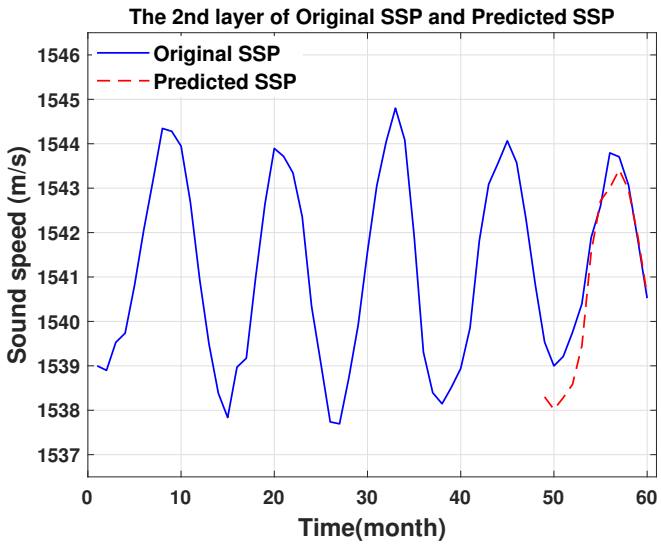
Table 5 compares the overall RMSE results of the four methods in predicting full-ocean depth SSP. H-LSTM achieves the most accurate predictions, with an RMSE of 0.5565 m/s.

Table 5. The comparison of RMSE of four prediction methods.

Area	168.5° E, 16.5° N	168.5° E, 16.5° N	168.5° E, 16.5° N	168.5° E, 16.5° N
Predict Time	2021.10	2021.10	2021.10	2021.10
Method	Mean value	Polynomial fitting	BP Neural Network	H-LSTM
Dataset	2017-10–2021.09	2019-10–2021.09	2017-10–2021.09	2017-10–2021.09
RMSE (m/s)	1.5959	0.9548	1.7861	0.5565

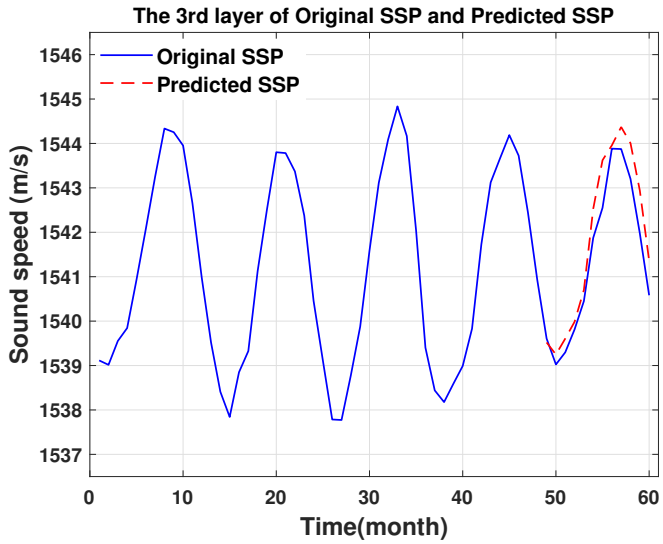
3.5. H-LSTM's Performance When Predicting Periodic Changes in SSPs

To assess the H-LSTM model's performance in predicting periodic changes in SSP, we utilized historical SSP data spanning from 2017 to 2020 as learning samples and conducted a 12-step prediction on future SSP data, forecasting SSPs for each month of the next year. We randomly selected the second layer (corresponding to a depth of 5 m), third layer (corresponding to a depth of 10 m), and fourth layer (corresponding to a depth of 20 m) between depth layers 1–58 to test the model's ability to accurately capture periodic changes in sound speed distribution. The comparisons of the periodic trends of the predicted SSP with the raw SSP data for the second, third, and fourth depth layers are shown in Figure 11.

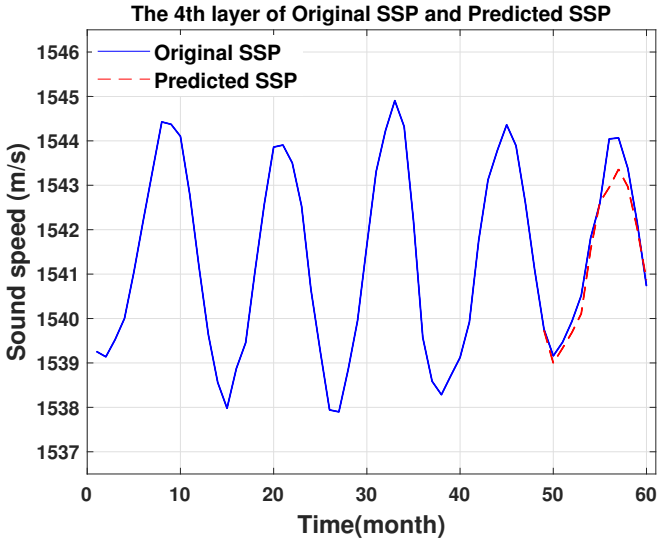


(**a**) The 2nd depth layer.

Figure 11. *Cont.*



(**b**) The 3rd depth layer.



(**c**) The 4th depth layer.

Figure 11. The comparison between the predicted trend and the original trend.

In the figures, the 60 solid blue lines depict real SSP data for the corresponding depth layers spanning 60 months from 2017 to 2021, with the first 48 months used for training and the last 12 for validation. The 12 red dashed lines represent predicted SSP data for the next 12 months at the respective depth layers, which are compared with the validation data. Our proposed H-LSTM method adeptly captures the periodic changes in SSP over time.

4. Conclusions

To meet the demands for accuracy and time efficiency in constructing underwater sound speed fields, we introduce an H-LSTM method for future full-ocean depth SSP prediction. With our SSP prediction method, sound speed distribution can be estimated without the need for on-site data measurements, significantly enhancing time efficiency.

To verify the feasibility and effectiveness of our model, we conducted experiments and established different state-of-the-art methods as the baseline. The experimental results demonstrate that the proposed H-LSTM method not only accurately predicts future full-ocean depth SSP but also effectively captures the periodic changes in SSP over time.

In future studies, we will focus on SSP prediction across multiple maritime regions, as well as the achievement of accurate prediction of oceanic SSPs in scenarios with limited sample data availability. This endeavor will entail the utilization of more intricate models and algorithms to effectively capture the dynamic variations in SSP distribution within complex marine environments, thereby furnishing more accurate data support for marine acoustic research.

Author Contributions: Conceptualization, J.L., W.H. and H.Z.; methodology, J.L. and W.H.; software, J.L. and W.H.; validation, J.L., S.L. and P.W.; formal analysis, J.L., W.H. and H.Z.; investigation, J.L., W.H. and H.Z.; resources, W.H. and H.Z.; writing—original draft preparation, J.L.; writing—review and editing, W.H., H.Z., S.L. and P.W.; funding acquisition, W.H. and H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Science Foundation of Shandong Province (ZR2023QF128), Laoshan Laboratory (LSKJ202205104), China Postdoctoral Science Foundation (2022M722990), Qingdao Postdoctoral Science Foundation (QDBSH20220202061), National Natural Science Foundation of China (NSFC:62271459), National Defense Science and Technology Innovation Special Zone Project: Marine Science and Technology Collaborative Innovation Center (22-05-CXZX-04-01-02), and the Fundamental Research Funds for the Central Universities, Ocean University of China (202313036).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Argo data were made available by the China Argo real-time data center on the web http://www.argo.org.cn/index.php?m=content&c=index&a=lists&catid=27 (accessed on 1 June 2024). The source codes presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Erol-Kantarci, M.; Mouftah, H.T.; Oktug, S. A survey of architectures and localization techniques for underwater acoustic sensor networks. *IEEE Commun. Surv. Tutor.* **2011**, *13*, 487–502. [CrossRef]
2. Liu, B.; Huang, Y.; Chen, W.; Lei, J. *Principles of Underwater Acoustics (In Chinese)*; China Science Publishing and Media Ltd. (CSPM): Beijing, China, 2019.
3. Zhang, S.; Xu, X.; Xu, D.; Long, K.; Shen, C.; Tian, C. The design and calibration of a low-cost underwater sound velocity profiler. *Front. Mar. Sci.* **2022**, *9*, 996299. [CrossRef]
4. GmbH, S.S.T. Sea & Sun Technology CTD Probes. 2023. Available online: https://www.sea-sun-tech.com/ (accessed on 1 June 2024).
5. Williams, A. CTD (conductivity, temperature, depth) profiler. In *Encyclopedia of Ocean Sciences: Measurement Techniques, Sensors and Platforms*; Steele, J.H., Thorpe, S.A., Turekian, K.K., Eds.; Elsevier: Boston, MA, USA, 2009; pp. 25–34.
6. Tsurumi-Seiki Co., Ltd. eXpendable Conductivity, Temperature and Depth Product Code: XCTD. 2023. Available online: https://tsurumi-seiki.co.jp/en/product/e-sku-2/ (accessed on 1 June 2024).
7. Huang, W.; Lu, J.; Li, S.; Xu, T.; Wang, J.; Zhang, H. Fast Estimation of Full Depth Sound Speed Profile Based on Partial Prior Information. In Proceedings of the 2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT), Qingdao, China, 21–24 July 2023; pp. 479–484. [CrossRef]
8. Tolstoy, A.; Diachok, O.; Frazer, L. Acoustic tomography via matched field processing. *J. Acoust. Soc. Am.* **1991**, *89*, 1119–1127. [CrossRef]
9. Taroudakis, M.I.; Markaki, M.G. Matched field ocean acoustic tomography using genetic algorithms. In *Acoustical Imaging*; Springer: Berlin/Heidelberg, Germany, 1995; pp. 601–606.
10. Choo, Y.; Seong, W. Compressive sound speed profile inversion using beamforming results. *Remote Sens.* **2018**, *10*, 704. [CrossRef]
11. Bianco, M.; Gerstoft, P. Compressive Acoustic Sound Speed Profile Estimation. *J. Acoust. Soc. Am.* **2016**, *139*, EL90–EL94. [CrossRef] [PubMed]
12. Li, Q.; Shi, J.; Zhenglin, L.; Yu, L.; Zhang, K. Acoustic sound speed profile inversion based on orthogonal matching pursuit. *Acta Oceanol. Sin.* **2019**, *38*, 149–157. [CrossRef]
13. Huang, W.; Li, D.; Jiang, P. Underwater Sound Speed Inversion by Joint Artificial Neural Network and Ray Theory. In Proceedings of the Thirteenth ACM International Conference on Underwater Networks & Systems (WUWNet'18), Shenzhen, China, 3–5 December 2018; pp. 1–8. [CrossRef]

14. Stephan, Y.; Thiria, S.; Badran, F. Inverting tomographic data with neural nets. In Proceedings of the 'Challenges of Our Changing Global Environment', San Diego, CA, USA, 9–12 October 1995; Conference Proceedings; OCEANS'95 MTS/IEEE; Volume 3, pp. 1501–1504. [CrossRef]
15. Munk, W.; Wunsch, C. Ocean acoustic tomography: A scheme for large scale monitoring. *Deep Sea Res. Part A. Oceanogr. Res. Pap.* **1979**, *26*, 123–161. [CrossRef]
16. Munk, W.; Wunsch, C. Ocean acoustic tomography: Rays and modes. *Rev. Geophys.* **1983**, *21*, 777–793. [CrossRef]
17. Taroudakis, M.I.; Markaki, M.G. On the use of matched-field processing and hybrid algorithms for vertical slice tomography. *J. Acoust. Soc. Am.* **1997**, *102*, 885. [CrossRef]
18. Yu, Y.; Li, Z.; He, L. Matched-field inversion of sound speed profile in shallow water using a parallel genetic algorithm. *Chin. J. Oceanol. Limnol.* **2010**, *28*, 1080–1085. [CrossRef]
19. Zhang, Z. A Study on Inversion for Sound Speed Profile in Shallow Water. P.h.D Thesis, Northwestern Polytechnical University, Xi'an, China, 2002. (In Chinese)
20. Liu, Y.; Chen, Y.; Meng, Z.; Chen, W. Performance of single empirical orthogonal function regression method in global sound speed profile inversion and sound field prediction. *Appl. Ocean Res.* **2023**, *136*, 103598. [CrossRef]
21. Dai, M.; Li, Y.; Ye, J.; Yang, K. An improved particle filtering technique for source localization and sound speed field inversion in shallow water. *IEEE Access* **2020**, *8*, 177921–177931. [CrossRef]
22. Zhang, W.; Yang, S.E.; Huang, Y.W.; Li, L. Inversion of sound speed profile in shallow water with irregular seabed. In Proceedings of the Advances in Ocean Acoustics: Proceedings of the 3rd International Conference on Ocean Acoustics (OA2012), Beijing, China, 21–25 May 2012; Volume 1495, pp. 392–399. [CrossRef]
23. Zhang, W. Inversion of Sound Speed Profile in Three-Dimensional Shallow Water. P.h.D Thesis, Harbin Engineering University, Harbin, China, 2013; Chapter 2. (In Chinese)
24. Zhang, L.; Liu, Y.; Liu, Y.; Chen, G.; Li, M. Modeling of Time-Varying Characteristics of Deep-Sea Sound Velocity Profile Based on Layered-EOF. *Coast. Eng.* **2022**, *41*, 209–222.
25. Huang, W.; Liu, M.; Li, D.; Yin, F.; Chen, H.; Zhou, J.; Xu, H. Collaborating Ray Tracing and AI Model for AUV-Assisted 3-D Underwater Sound-Speed Inversion. *IEEE J. Ocean. Eng.* **2021**, *46*, 1372–1390. [CrossRef]
26. Li, Q.; Li, H.; Cao, S.; Yan, X.; Ma, Z. Inversion of the Full-depth SSP based on Remote Sensing Data and Surface Sound Speed. *Mar. Bull.* **2022**, *44*, 84–94.
27. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
28. Zhang, C.; Liu, Z.; Lu, S.; Li, H.; Sun, C.; Wu, X. *User Manual (3rd Version) of GDCSM Argo Gridded Data Set*; China Argo Real-time Data Center: Tianjin, China, 2021.
29. Liu, F.; Tuo, J.I.; Zhang, Q. Sound Speed Profile Inversion Based on Mode Signal and Polynomial Fitting. *Acta Armamentarii* **2019**, *40*, 2283–2295.
30. Yu, X.; Xu, T.; Wang, J. Sound Velocity Profile Prediction Method Based on RBF Neural Network. In *China Satellite Navigation Conference (CSNC)*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 1–8.