



Article Reconstruction of the Sound Speed Profile in Typical Sea Areas Based on the Single Empirical Orthogonal Function Regression Method

Wen Chen ¹, Kaijun Ren ¹, Yongchui Zhang ^{1,*}, Yuyao Liu ¹, Yu Chen ^{1,*}, Lina Ma ¹ and Silin Chen ²

- ¹ College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China; chenwen_hql@126.com (W.C.)
- ² School of Journalism, Yunnan University, Kunming 650000, China
- * Correspondence: zyc@nudt.edu.cn (Y.Z.); chenyu@nudt.edu.cn (Y.C.)

Abstract: The sound speed profile (SSP) is a necessary prerequisite for acoustic field computation and underwater target localization and monitoring. Due to the dynamic nature of the ocean, the reconstruction of SSPs with surface characteristics is a big challenge. In this study, the Single Empirical Orthogonal Function Regression (sEOF-R) method is employed to establish the regression relationship between the surface parameters and the sound speed anomaly profile (SSAP) in three typical sea areas, namely the equator, Kuroshio Extension (KE), and Northeast Pacific. Based on the established regression relationship and the surface parameters, the underwater SSP is reconstructed. Results show that the reconstruction effects in the three areas show the best performance in the Northeast Pacific, followed by the equator and finally the KE. The quantitative analysis suggests that the local sea level anomaly (SLA) plays the dominant role in influencing the reconstruction effect, followed by the sea surface temperature anomaly (SSTA). Further analysis demonstrates that the sEOF-R method is limited in time-varying and space-varying areas. The SSP reconstructed from the sea surface information in this study is useful for the inversion of the underwater structures.



1. Introduction

The low energy loss of sound waves in water makes them the main approach for underwater target detection and location. However, influenced by the inhomogeneous water body, sound waves propagate through refraction and reflection and produce complex paths. In underwater acoustic engineering, such inhomogeneity is characterized by the range-dependent sound speed profiles (SSPs), which are derived from the temperature (T) and salinity (S) profiles. Owing to the oceanic dynamic characteristics, however, the T and S profiles and SSPs are all highly temporally and spatially dependent, which will significantly change the arrival structures (such as time delay, amplitude, angle) of sound, thus decreasing the performance of underwater localization and detection algorithms. The different effective working distances of the sonar system in the morning and afternoon on the same day, known as the "afternoon effect", is a classic example demonstrating the high dependence of hydroacoustic equipment on the marine environment. Thus, determining how to estimate the sound speed profile (SSP) accurately and in real time has been a difficult research topic in recent decades [1–6].

There are many ways to obtain the SSP, such as in situ observation, oceanic environmental numerical simulation, and inversion methods using sea surface parameters. For the in situ observation method, observational instruments, such as Argo buoys, expendable conductivity temperature depth (XCTD) [7], expendable bathythermograph (XBT) [8], and sound velocity profile (SVP) [9], are applied to obtain the hydrological parameters such as



Citation: Chen, W.; Ren, K.; Zhang, Y.; Liu, Y.; Chen, Y.; Ma, L.; Chen, S. Reconstruction of the Sound Speed Profile in Typical Sea Areas Based on the Single Empirical Orthogonal Function Regression Method. *J. Mar. Sci. Eng.* 2023, *11*, 841. https:// doi.org/10.3390/jmse11040841

Academic Editor: Alberto Ribotti

Received: 14 March 2023 Revised: 11 April 2023 Accepted: 14 April 2023 Published: 15 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). T, S, and pressure. Although this method can obtain the highest-accuracy SSP, conducting in situ experiments is difficult because they are resource- and time-consuming endeavors and are performed only by a small number of communities [10–12].

Oceanic numerical simulation refers to the process of discretizing and solving seven ocean dynamic equations with the help of boundary conditions and part of ocean observation data to obtain the three-dimensional distribution of hydrological elements. The method can cover the global oceans with a spatial resolution reaching up to $3 \text{ km} \times 3 \text{ km}$ [13], and the reliability of the simulated data can be significantly improved by the numerical assimilation technology [14–17]. Up to now, the simulation data from numerical models have been widely used in the analysis of various physical phenomena in oceanography. However, the expensive consumption in calculation and time remains a challenge.

Compared to the sparse in situ observation profiles, the sea surface parameters (e.g., sea surface temperature, sea surface height) observed by satellites with high spatialtemporal resolution and temporal refresh rate are easy to obtain. In addition, the effective detection distances of the sonar array are less than 50 km, demanding a 5 km resolution of SSPs. However, the spatial resolution of a quality-controlled SSP is generally 100 km \times 100 km [18], which cannot meet the requirements for the SSP resolution of a sonar system. In conjunction with the refinement of satellite datasets, inversion methods involving statistics and regression can meet the accuracy requirements and have gained attention and development [19,20]. The core of this technique is to build the mapping relationships between surface and underwater parameters. The Empirical Orthogonal Function (EOF) is a frequently used method to analyze the structural features in matrix data and extract the main data features. Usually, only the first two- or three-order EOFs can accurately represent the SSP in the measurement area. Park and Kennedy [21] proved that at most the first five-order EOFs could accurately represent the SSP even in the case of significant differences from the average. Benefiting from such an ability of the EOF, Carnes et al. [22] fulfilled the streamlining and compression of historical T and S profiles over 33,000.

Once the regression relationships between sea surface parameters and the coefficients of the first several-order EOFs corresponding to the historical underwater SSP dataset are established, the SSPs with the corresponding spatial resolution can then be estimated through the sea surface parameters with high spatial and temporal resolution. This kind of strategy was employed in many works. The Navy's Modular Ocean Data Assimilation System (MODAS) [19] used multivariate regression between historical observations of the sea surface height (SSH) and sea surface temperature (SST) and T and S at defined depths to compute the T and then estimate S from historical T-S regressions. It can produce a rapid estimation of present and near-term ocean conditions, even in situations where little or no in situ data are available [19]. The Improved Synthetic Ocean Profile (ISOP) method constructs the dynamic interior layer using covariances of SST and SSH with coupled EOFs of T and S, as well as coupled EOFs of vertical gradients of T and S in the upper 1000 m [21]. Compared to the MODAS, ISOP shows better performance in terms of 48 h sonic layer depth and below-layer gradient forecast. Yan et al. [23] presented a density incorporation tool of a least square multivariate empirical orthogonal function algorithm, which can retrieve interior T/S from the ocean surface and outperform machine learning algorithms in retrieving complicated T/S structures. Chen et al. [6] established the regression relationships between sea level anomaly (SLA) and sea surface temperature anomaly (SSTA) and the historical sound speed anomaly profiles (SSAPs) using the Empirical Orthogonal Function Regression (EOF-R) method and realized the reconstruction of SSPs on a global scale. Because there are two input variables and only one output (SSAP), thus this method is called the Single Empirical Orthogonal Function Regression (sEOF-R) method. In addition, the statistical relationships between reconstruction error and the local SSTA and the eddy kinetic energy are discussed.

Though the sEOF-R method has been widely employed in worldwide SSP reconstruction, in areas with abundant dynamic systems, such as Kuroshio Extension (KE), the actual effect of applying the sEOF-R method remains to be verified and discussed [24]. In addition, the reconstruction effect of the sEOF-R method in different layers needs to be analyzed and discussed. Furthermore, whether the reconstructed SSP can serve hydroacoustic equipment also needs to be verified.

To answer the above three questions, the underwater SSPs in three typical areas (the equator, Kuroshio Extension (KE), and Northwest Pacific) are reconstructed based on the sEOF-R method. The impacts of intensities of the SSTA and SLA on the performance of the sEOF-R method in the SSP reconstruction are also analyzed. An analysis is then made on the difference in sound propagation based on the original and reconstructed SSPs. Finally, instructions for the use of the sEOF-R method and suggestions for its improvement are offered.

The remaining content of the paper is outlined as follows: In Section 2, the input datasets are introduced first, followed by the establishment of the principle and application process of the sEOF-R method in reconstructing SSP. Section 3 includes the reconstructed results and the statistical relationship between the surface parameters and the root mean square error (RMSE) of SSPs. The comparative analysis of sound propagation characteristics is given in Section 4. Finally, we end in Section 5 with conclusions and recommendations.

2. Material and Methods

2.1. Oceanic Datasets

Global Ocean Argo Grid Dataset 2018 (BOA_Argo 2018) (http://www.argo.org.cn/, accessed on 11 January 2023), which integrates Argo T/S data from 2004 to 2017, provides monthly averaged gridded variables (S, T, Isothermal Layer Depth, Mixed Layer Depth and Composed Mixed Layer Depth) with a horizontal resolution of 1° at 58 depth levels (from 0 to 1975 m), including the 3D-grid T/S data [25]. The monthly T/S profiles in BOA_Argo 2018 dataset were employed to compute the corresponding monthly SSP through the Wilson sound speed empirical formula. Throughout the rest of the text, the SSPs derived from BOA_Argo 2018 will simply be called "Argo SSPs", and they are treated as the measured SSPs in this work. One part of them was used to establish the regression relationships between SLA, SSTA, and the coefficients of EOFs, the other part was used to verify the performance of the sEOF-R method in the SSP reconstruction.

Monthly 1/12° grid sea surface altimeter data and SST data were both provided by Copernicus Marine Environment Monitoring Service (CMEMS) (https://marine.copernicus.eu/, accessed on 11 January 2023). Among them, monthly mean sea level anomalies (SLAs) are referenced to the period from 1993 to 2012. Similar to BOA_Argo 2018 dataset, CMEMS provides the SST and SLA that also have direct/indirect usages in two aspects of this work. On the one hand, CMEMS provides the SST and SLA as the historical data in establishing the regression relationships. On the other hand, CMEMS provides SST and SLA to estimate the underwater SSP in the reconstruction section. Noteworthily, the SSTA is calculated by subtracting climatological SST from monthly SST instead of directly provided by CMEMS.

World Ocean Atlas 2018 (WOA18) (https://www.ncei.noaa.gov/archive/accession/ NCEI-WOA18, accessed on 11 January 2023) is a dataset of objectively analyzed onedegree-grid climatological field in situ T/S at standard depth levels for annual, seasonal, and monthly compositing periods for oceans in the world. Considering the monthly differences of SSP structures, the monthly compositing periods (2005–2017) of the $1^{\circ} \times 1^{\circ}$ grid dataset were employed in this study. WOA18 was used to provide the climatological SSPs involving two steps in this work. One was to obtain the sound speed anomaly profile (SSAP) by removing the climatological SSP from the Argo SSP. The other was to calculate the reconstructed SSP by adding climatological SSP to the SSAP estimated from the SLA and SSTA. In addition, WOA18 was used to estimate the climatological SST, which was then employed to calculate the SSTA by subtracting it from the monthly SST in CMEMS. See Section 2.2 for more details.

2.2. SSP Reconstruction Based on the sEOF-R Method

Regression is a statistical method that attempts to determine the strength and character of the relationship between one dependent variable and a series of other independent variables. Linear regression is the most common form of this technique. With the aid of the regression method, the sEOF-R method establishes the regression relationships between SLA and SSTA (independent variables) and the coefficients (dependent variables) of the first four-order EOFs of the SSAP matrix. According to the established regression relationships, the underwater SSP can be reconstructed by SLA and SSTA. Since sufficient historical data are required for applying the sEOF-R method, the input data from 2005 to 2017 were used. In addition, the selected areas were divided into $2^{\circ} \times 2^{\circ}$ grids, and then the SSPs were reconstructed monthly. As illustrated in Figure 1, there are nine Argo SSPs (black dots) and four climatological SSPs (red circles) in a $2^{\circ} \times 2^{\circ}$ grid. The climatological SSP is represented by the mean of the WOA18 SSPs located in four vertices of the grid.



Figure 1. Diagrammatic sketch of a $2^{\circ} \times 2^{\circ}$ grid. The hollow red circle represents the climatological SSP derived from T/S in WOA18, and the black dot represents the Argo SSP derived from T/S in BOA_Argo 2018 dataset.

The sEOF-R method reported in [6,24] was employed in SSP reconstruction. The process of the sEOF-R method is established in Figure 2. With the aid of the empirical formula of sound speed, the Argo SSP and climatological SSP can be derived from BOA_Argo 2018 and WOA18 datasets, respectively. Then, SSAPs can be obtained by subtracting climatological SSP from the corresponding Argo SSP. Applying the EOF method in the SSAP matrix, the first four-order EOFs and corresponding coefficients can be obtained. Simultaneously, SSTA is obtained by subtracting climatological SST from the real-time SST. SLA can be directly obtained in CMEMS. We can establish the regression relationship between SLA, SSTA, and coefficients of the first four-order EOFs and then store the regression relationship coefficients in the historical regression relationship library. Ultimately, the reconstructed SSAPs can be inferred by using SLA, SSTA, and the regression relationship library. The reconstructed SSP is obtained by superposing the climatological SSP.

The SSP is calculated by the Wilson sound speed empirical formula, using T, S, and static pressure. As mentioned in Section 2.1, the SSAP is computed by subtracting the corresponding climatological SSP from the Argo SSP. SSAPs are then arranged in the SSAP matrix Q in spatial and temporal order. The dimension of Q is $M \times N$, where M and N denote the depth layers and quantity of SSAPs, respectively. The EOF method is applied to Q in order to obtain the *m*-order empirical orthogonal function f_m (m = 1, ..., M) and its coefficient matrix λ . The mathematical formula for the EOF method is given by

$$R = Q \times Q^{H},$$

$$(R - \lambda I)F = 0,$$
(1)

where the superscript *H* denotes the conjugate transpose of the matrix, *R* denotes the covariance matrix of the SSAP matrix $Q, F = \begin{bmatrix} f_1 & \dots & f_M \end{bmatrix}$ denotes the empirical orthogonal matrix, $\lambda = diag(\lambda_1, \dots, \lambda_M)$ denotes the coefficient matrix and $\lambda_m(m = 1, \dots, M)$ denotes the empirical orthogonal coefficient for the *m*-order EOF. The SSAPs (represented by ΔSSP) are described by the linear combination of the first four-order EOFs, since the weight of the coefficients of the first four-order EOFs is greater than 92%.

$$\Delta SSP = \alpha_0 + \sum_{m=1}^4 \alpha_m f_m \tag{2}$$

where α_m , denotes the coefficient of the m-order EOF and α_0 denotes the constant coefficient. The regression relationship between sea surface parameters (SLA, SSTA) and α_m is then established through linear regression analysis:

$$\alpha_m(t) = b_0 + b_1 SLA(t) + b_2 SSTA(t) + b_3 SLA(t) \times SSTA(t), \ m = 0, 1, \dots, 4,$$
(3)

where $\mathbf{b} = \begin{bmatrix} b_0 & b_1 & b_2 & b_3 \end{bmatrix}$ is the library of regression relationship coefficients and $t \in \begin{bmatrix} 1, 117 \end{bmatrix}$ denotes the number of SSAPs. The reconstructed SSP (*SSP*_{res}) can be estimated by adding the reconstructed SSAP (ΔSSP) to the climatological SSP (*SSP*)



Figure 2. The process flow diagram for the SSP reconstruction. SSAP, sound speed anomaly profile; SLA, sea level anomaly; SSTA, sea surface temperature anomaly.

2.3. Reconstruction Errors

To demonstrate the regional characteristics of three areas, SLA_I and $SSTA_I$, which are used to quantitatively depict the monthly fluctuation intensity of SLA and SSTA, are introduced. Their mathematical forms are shown as follows:

$$SLA_{I}(mon) = \sqrt{\frac{1}{N} \sum_{n=1}^{N} SLA^{2}(n, mon)},$$
(5)

$$SSTA_{I}(mon) = \sqrt{\frac{1}{N} \sum_{n=1}^{N} SSTA^{2}(n, mon)},$$
(6)

where *mon* denotes the month and *N* is the total amount of SSPs in a $2^{\circ} \times 2^{\circ}$ grid in a month during the study time. As mentioned previously, the total data for thirteen years from 2005 to 2017 are selected, leading to the equation of N = 117 in a $2^{\circ} \times 2^{\circ}$ grid.

In order to quantitatively evaluate the effects of the sEOF-R method in the SSP reconstruction, *RMSE* and *RMSE*_{depth} are employed as evaluation metrics. *RMSE* represents the root mean square error between the reconstructed SSPs and the corresponding Argo SSPs which represent the real SSPs in a $2^{\circ} \times 2^{\circ}$ grid. The mathematical formula is given by

$$RMSE(mon) = \sqrt{\frac{1}{M \times N} \sum_{n=1}^{N} \sum_{m=1}^{M} (c_{res}(n, m, mon) - c_{Argo}(n, m, mon))^2}, \quad (7)$$

where *mon* denotes month, *N* denotes the total number of SSPs contained in a $2^{\circ} \times 2^{\circ}$ grid, *M* denotes the total number of layers, and c_{res} and c_{Argo} denote the reconstructed and Argo sound speed values in m/s. *RMSE(mon)* represents the monthly reconstruction error of SSPs in a $2^{\circ} \times 2^{\circ}$ grid in *mon* month. It considers the errors between the reconstructed and Argo SSPs at all depths in a $2^{\circ} \times 2^{\circ}$ grid; thus, the reconstruction effects in depth direction are averaged and cannot be represented.

To show the reconstruction effect in depth, the reconstruction error ($RMSE_{depth}(m, mon)$) in *mon* month and *m* layer is introduced. Unlike *RMSE*, it considers the root mean square error between the reconstructed SSPs and the corresponding Argo SSPs layer by layer, thus reserving the structural features vertically. The mathematical formula is written as

$$RMSE_{depth}(m,mon) = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (c_{res}(n,m,mon) - c_{Argo}(n,m,mon))^2}, \qquad (8)$$

where $c_{res}(n, m, mon)$ and $c_{Argo}(n, m, mon)$ denote the sound speed values (unit: m/s) in *mon* month, *n* SSP, and *m* layer in reconstructed and Argo SSPs, respectively. Thus, the vertical errors of the reconstructed and Argo SSPs are obtained.

2.4. Study Areas

In order to verify the performance of the sEOF-R method, three typical ocean areas marked as A, B, and C in Figure 3 were selected for the analysis. Area A (2.5° N–12.5° N, 144.5° E–164.5° E), Area B (28.5° N–38.5° N, 144.5° E–164.5° E), and Area C (28.5° N–38.5° N, 131.5° W–151.5° W) are separately located in the north of the equator, KE, and Northeast Pacific Ocean, respectively. The reasons for the selection of these areas are as follows:

Area A: The SSP structure is stable, with the thin mixed layer (ML) existing throughout the year. The *SLA*₁ range is between 0.04 m and 0.16 m; the *SSTA*₁ range is between 0.23 °C and 1.22 °C

Area B: The oceanic fronts and mesoscale eddies are common phenomena all year round in this area. Thus, the fluctuations of SLA and SSTA are both intense, with the corresponding SLA_I ranging between 0.06 m and 0.44 m and $SSTA_I$ ranging between 0.52 °C and 2.54 °C. Area C: As Area C and Area B are located in the same latitude range, changes in the SSP structure and ML in these two areas follow the same rules. However, the *SLA* range in Area C is much smaller than that in Area B, with the *SLA_I* ranging between 0.035 m and 0.12 m and *SSTA_I* ranging between 0.46 °C and 2.9 °C.



Figure 3. The monthly average SLA and SSTA in July are shown in the left and right panels, respectively. Areas A, B, and C which are located in the equator region, KE, and Northeast Pacific Ocean are marked as dashed rectangles.

3. Results

3.1. The SSP Reconstruction

The depth sampling interval of WOA18 is different from that of Argo. Thus, interpolation processing is necessary, with the WOA18 SSP and Argo SSP obtained with a depth interval of 5 m after cubic spline interpolation. Considering the depth limitation in the WOA18 dataset, SSPs with a maximum depth of 1500 m are selected for the research. In addition, to avoid calculation errors due to the structural differences of SSPs in different months, the SSPs need to be reconstructed monthly.

The reconstructed SSPs in January, April, July, and October during thirteen years in Areas A, B, and C are shown in Figure 4, where the solid black line denotes Argo SSPs and the red dotted line denotes the reconstructed SSPs. From the left column of Figure 4, it can be found that a high structural consistency exists in both the Argo SSP and the reconstructed SSP, mainly due to the stable T/S structure in the equator area.



Figure 4. The SSPs in January, April, July, and October in Areas A (**a**–**d**), B (**e**–**h**), and C (**i**–**l**), with the black line and red line representing the Argo SSPs and reconstructed SSPs, respectively. The insets show the SSPs in the depth ranging from 0 m to 100 m.

Area B is located in KE, and the reconstructed SSP here is shown in Figure 4 (middle column). It reveals the underlying law for the typical periodic change of SSPs in the midlatitude areas; that is, the thickness of the ML of SSP changes significantly with the seasons, with gradual thinning and disappearing in summer and gradual appearing and thickening in winter. It can be easily found in the middle column of Figure 4 that a sound structural consistency exists between the reconstructed SSP and the Argo SSPs. A comparison of data in the left and middle columns shows that there is an obvious increase in the dispersion of the Argo SSP in Area B, which means a significant performance degradation of the sEOF-R method in the dynamic areas. This phenomenon is mainly caused by the increase in SLA_I and $SSTA_I$ in KE.

The reconstructed results in Area C are shown in Figure 4 (right column). The depth of the SOFAR axis can be inferred accurately, and high structural consistency in the thermocline and deep-sea isothermal layer can be found between the reconstructed SSP and Argo SSP. As described at the beginning of Section 3, the $SSTA_I$ in Area C is stronger than that in Area A. Thus, the reconstruction error in Area C is bigger than that in Area A in the ML.

3.2. Effect Analysis of the sEOF-R Method in SSP Reconstruction

In Section 3.1, it can be concluded that the performances of the sEOF-R method in the SSP reconstruction are related to the local SLA_I and $SSTA_I$. To quantitatively analyze the impact of the local SLA_I and $SSTA_I$ on the performance of reconstruction, we establish two linear relationships between RMSE and SLA_I and $SSTA_I$, respectively, which are written as

$$RMSE(mon) = p_{SLA}^{(1)} \times SLA_I(mon) + p_{SLA}^{(2)}$$
(9)

$$RMSE(mon) = p_{SSTA}^{(1)} \times SSTA_I(mon) + p_{SSTA}^{(2)}$$
(10)

where $p^{(1)}$ and $p^{(2)}$ are calculated by the least squares method (LSM) method, and the coefficients $p_{SLA}^{(1)}$ in s⁻¹ and $p_{SSTA}^{(1)}$ in ms⁻¹ °C⁻¹ denote the sensitivity of reconstruction error to SLA_I and $SSTA_I$, respectively. A bigger $p^{(1)}$ means a higher sensitivity.

The RMSE of reconstructed SSPs in Areas A, B, and C in January, April, July, and October in the period of 2005 to 2017 are shown in Figure 5, where the red line calculated by Equation (9) represents the linear relationship between RMSE and SLA_I . To shorten the length of the article, only the results for January, April, July and October, which represent the spring, summer, autumn, and winter seasons, are shown. It can be found that the sensitivity ($p_{SLA}^{(1)}$) of RMSE to SLA_I is positive in most months except for February (not shown) to April. In Area A, the minimum $p_{SLA}^{(1)}$ (-2.31 s^{-1}) occurs in March, and the maximum (7.73 s⁻¹) occurs in December. $p_{SLA}^{(1)}$ becomes higher in winter but lower in summer. Compared to the SLA_I , the relationship between RMSE and $SSTA_I$ in Area A is less significant, as shown in Figure 6.

In Area B, the scattered RMSE points are well fitted by the red line as illustrated in the middle columns of Figures 5 and 6, which means there is a significant relationship between RMSE and SLA_I ($SSTA_I$). Among the three areas, it is the most significant area, with $p_{SLA}^{(1)}$ ranging from 12.07 (s⁻¹) to 16.15 (s⁻¹). Similarly, $p_{SSTA}^{(1)}$ is generally higher in Area B, as shown in Figure 6.

As illustrated in Figure 5 (left column), the relationship between the SLA_I and RMSE in Area C is not straightforward. However, the linear relationships between $SSTA_I$ and RMSE are prominent (Figure 6), with the $p_{SSTA}^{(1)}$ being positive and ranging from 0.17 (ms⁻¹ °C⁻¹) to 1.27 (ms⁻¹ °C⁻¹).



Figure 5. The relationship between *SLA*_{*I*} (m) and RMSE (m/s) in Areas A (**a**–**d**), B (**e**–**h**), and C (**i**–**l**) in January, April, July, and October. The blue circles denote the scattered RMSE points in grids.



Figure 6. The relationship between *SSTA*_{*I*} (°C) and RMSE (m/s) in Areas A (**a**–**d**), B (**e**–**h**), and C (**i**–**l**) in January, April, July, and October. The blue circles denote the scattered RMSE points in grids.

To facilitate the observation, $p_{SLA}^{(1)}$ and $p_{SSTA}^{(1)}$ in Areas A, B, and C are all recorded in Table 1, with the superscript A, B, and C representing Areas A, B, and C, respectively. Through the comparison of the SLA and SSTA sensitivity ($p^{(1)}$) among Areas A, B, and C, it can be found that the SLA_I and $SSTA_I$ both contribute to the RMSE, with the SLA_I being the more important influential factor. The enhancement of the SLA_I and $SSTA_I$ will lead to a clearer linear relationship between the sea surface parameters and RMSE. On the contrary, the decrease in the SLA_I and $SSTA_I$ will weaken the linear relationship.

Month	1	2	3	4	5	6	7	8	9	10	11	12
p_{SLA}^A	3.27	-0.26	-2.31	-0.10	3.98	4.26	4.68	6.12	5.03	1.62	5.24	7.73
p^{A}_{SSTA}	0.3	0.82	-0.81	-0.7	2.93	0.49	1.94	-1.36	0.71	-0.42	-0.19	-0.56
p_{SLA}^B	14.36	14.07	14.21	16.15	16.11	16	12.93	12.07	14.42	13.2	14.44	14.37
p^B_{SSTA}	2.15	2.4	2.23	1.04	-0.19	-0.09	2.55	2.92	1.81	2.64	2.22	2.85
p_{SLA}^C	-2.17	4.7	6.34	-7.11	1.41	12.44	8.82	4.76	16.48	3.41	5.37	6.65
p_{SSTA}^{C}	0.17	0.43	0.30	0.47	0.45	0.38	0.32	0.36	1.27	0.25	0.42	0.36

Table 1. Sensitivity of RMSE to *SLA_I* and *SSTA_I*.

Furthermore, the linear relationships in Table 1 are verified by the significance test; the associated *p*-values are shown in Table 2. In Area A, the p_{SLA}^A in January to April, October, and December and the p_{SSTA}^A in January to July as well as October to December do not pass the 95% significance test. In Area B, the p_{SSTA}^B values in April to June do not pass the 95% significance test. In Area C, the p_{SLA}^C values in January to May do not pass the 95% significance test. It can be found that only the relationships between SLA and RMSE in Area B and between SSTA and RMSE in Area C pass the 95% significance test. The results in Table 2 support that the reconstruction error will increase with local SLA (SSTA) in Area B (Area C), which indicates that the s-EOF method is limited in dynamically active regions.

Table 2. The <i>p</i> -values in significance t	est.
--	------

Month		1	2	3	4	5	6	7	8	9	10	11	12
A A	SLA	0.33	0.92	0.21	0.95	0.01	0.00	0.00	0.00	0.00	0.54	0.05	0.07
Area A	SSTA	0.93	0.06	0.18	0.17	0.24	0.27	0.07	0.02	0.00	0.43	0.87	0.44
	SLA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Area B	SSTA	0.00	0.00	0.00	0.06	0.79	0.89	0.00	0.00	0.00	0.00	0.00	0.00
Area C	SLA	0.51	0.27	0.16	0.26	0.84	0.03	0.01	0.04	0.00	0.04	0.03	0.02
	SSTA	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00

3.3. The Vertical Characteristics of RMSE

In order to discover the underwater performances of the s-EOF method, the vertical RMSE is obtained. The $RMSE_{depth}$ and the first one-order EOF from January to December in Areas A, B, and C are shown in Figure 7, marked with red and blue solid lines, respectively. In addition, the impacts of seasonal influence on the structure of the $RMSE_{depth}$ in Areas A, B, and C are also considered. Due to the stable T/S structure of the seawater in Area A, the maximum response depth (MRD) and the maximum error depth of $RMSE_{depth}$ in the twelve subfigures are highly consistent (Figure 7, upper panel) and are both at a depth range from 200 m to 250 m.

The influence of SLA_I and $SSTA_I$ on SSP reconstruction can be discovered by comparing the $RMSE_{depth}$ within the ML in Figure 7. As described at the beginning of Section 3, the range of SLA_I in Area C is similar to that in Area A, but the $SSTA_I$ in Area C is about twice stronger than that in Area A. Therefore, the RMSE at the sea surface in Figure 7 (bottom panel) is bigger than that in Figure 7 (upper panel), which is consistent with the results summarized in Table 1. Similarly, the range of $SSTA_I$ in Area C is similar to that in Area B, but the SLA_I in Area B is 4 times as strong as that in Area C, as illustrated at the



beginning of Section 3. Thus, the RMSE in depths ranging from 0 m to 100 m in Figure 7 (middle panel) is bigger than that in Figure 7 (bottom panel).

Figure 7. The vertical structure of the first one-order EOF and $RMSE_{depth}$ of SSP in Areas A (A1–A12), B (B1–B12), and C (C1–C12) with the blue and red lines representing the first one-order EOF and $RMSE_{depth}$ of the SSP.

Prior to the exploration of the reasons for the above phenomena, the fact that the structural similarity of SSPs determines the effect of SSP reconstruction needs to be clarified. The SSP dataset with the higher structural similarity and simpler type can generate the first four-order EOFs with a higher cumulative contribution rate, which means more features of SSPs are concentrated in the first four-order EOFs, resulting in a much better reconstruction effect under this circumstance. Simultaneously, it is known that the sound speed is mainly affected by the T, and SLA_I and $SSTA_I$ can indirectly and directly reflect the magnitude of temperature fluctuations. So, SLA_I and $SSTA_I$ are able to affect the reconstruction effect of the SSP indirectly.

Next, an analysis is conducted on the formation causes of the $RMSE_{depth}$ by layers. In the ML, the water temperature change can be reflected directly by SLA_I and $SSTA_I$ due to their close relationship with the sea surface. Thus, the higher SLA_I and $SSTA_I$ in Area B will inevitably lead to the complexity of SSP types and the declining cumulative contribution rate of the first four-order EOFs, which will result in the degradation in SSP reconstruction in Figure 7 (middle panel). In the isothermal layer in any of the sea areas where the temperature of the seawater is constant, there is a fixed linear relationship between sound speed and depth. Therefore, the structure of the SSP dataset with a single type here is always stable, so the cumulative contribution rate is always high, and the effect of reconstruction is the best as shown in Figure 7. In the thermocline, which is far from the sea surface, the water temperature change is mainly determined by the internal water mass, instead of the SLA and SSTA, thus weakening the correlation between sea surface parameters and SSP. In other words, the SSP change in the thermocline can be reflected not only by the water mass, but also by the sea surface parameters, which will reduce the cumulative contribution rate of the first four-order EOFs. For these two reasons, the performance of the sEOF-R method is the worst in the thermocline, as shown in Figure 7.

Ultimately, the ranges and median of the RMSE and $RMSE_{depth}$ are summarized in Table 3. Based on the comparison of these data, it can be found that the effect of reconstruction is the worst in Area B and the best in Area C. Furthermore, by comparing the $RMSE_{depth}$ in Figure 7, we discovered remarkable differences in MRD, which ranges from 200 m to 250 m, from 500 m to 700 m, and from 180 m to 200 m in Areas A, B, and C, respectively. These differences may be caused by the variances in the thermohaline structure, but the MRD of the first one-order EOF and the depth of the maximum $RMSE_{depth}$ are very close in a comparison of the first one-order EOF and corresponding $RMSE_{depth}$ curves.

Table 3. The RMSE and *RMSE*_{depth} in Areas A, B, and C.

Areas	А	В	С
$RMSE_{depth}$ (m/s)	0.15~12.39	0.15~14.03	$0.11 \sim 10.08$
RMSE(m/s)	1.17~3.01	1.21~7.32	1.06~3.33
Median of $RMSE_{depth}$ (m/s)	0.58	1.59	0.49
Median of RMSE (m/s)	1.73	3.68	1.39

4. Discussion

SSP reconstruction has extensive application prospects. In this section, the SSP reconstruction effect is evaluated according to the sound field calculation by using BELL-HOP [26], a tool for sound propagation calculation. Meanwhile, an analysis is conducted on the changes in acoustic propagation characteristics caused by SSP reconstruction errors. To eliminate the influence of the seabed, the topography is set to be flat with a depth of 1500 m. The type of bottom boundary condition is set as ACOUSTO-ELASTIC half-space with a density of 1.6 g/cm^3 .

The Argo and reconstructed SSPs in Areas A, B, and C are shown in Figure 8. In Area A, they are highly consistent in the depth of the SOFAR axis, the gradient of the SSP, and the sound speed at the surface. Due to Area A being located on the equator, the sound speeds at the surface and SOFAR axis are both higher with values of ~1530 m/s and ~1484 m/s. From the acoustic perspective, significant differences occur in the mixed layer depth, with depths of ~40 m and ~60 m in Argo and reconstructed SSPs, respectively. Compared with SSPs in Area B and C, the sound speeds in the isothermal layer in Area A are highest with the minimum vertical gradient of the SSP.



Figure 8. The Argo SSP and the reconstructed SSP in (**a**) Area A, (**b**) Area B, and (**c**) Area C, marked with black and red dashed lines, respectively.

In Area B, the obvious difference between the Argo and reconstructed SSPs occurs near the SOFAR axis. The depth difference in the SOFAR axis can reach up to \sim 100 m, and the sound speed difference can be up to \sim 5 m/s. In the thermocline, the Argo SSP exhibits a uniform negative vertical gradient, while the change in the reconstructed SSP is relatively moderate, resulting in a higher sound speed at 500 m in the reconstructed SSP. In Area C, the maximum reconstruction errors occur at the sea surface (~7 m/s) and are followed by the ML as shown in Figure 8c, and errors decrease with depth. Despite the values of sound speed in the reconstructed SSP being inaccurate, the reconstructed SSP precisely portrays changing trends in the ML, thermocline, and isothermal layer.

Figure 9 presents the transmission loss (TL, unit: dB) results in Areas A, B, and C. Figure 9a,b exhibit the sound fields with the source depth (SD) set at 20 m and frequency of 100 Hz under the Argo SSP and reconstructed SSP in Area A. In Figure 9a,b, the energy obtained from the source falls into two categories. One part is trapped within the ML to form the surface channel, and the remaining part enters the thermocline to form the deep-sea channel. Due to the acoustic absorption of the seabed, the convergence zones (CZs) of the deep-sea channel gradually disappear with the increase in propagation distance. As shown in Figure 9c, the TL at the receive depth (RD) of 50 m calculated by the reconstructed SSP is slightly lower than that calculated by the Argo SSP. In addition, the two TL curves manifest an obvious structural difference. Specifically, there are two types of CZs as illustrated in red with an interval of 13 km, marked as Z1 and Z2, respectively. Z1 represents the convergence zone (CZ) generated from the deep-sea channel, while Z2 represents CZ generated from the surface duct. As shown in Figure 9c, there is high consistency between Z1 and CZ from Argo SSP. Z2 is generated from the bigger negative gradient of reconstructed SSP in the ML.

There are oceanic fronts and mesoscale eddies all year round in Area B, resulting in more structurally diverse underwater SSPs. This means the weight (95.83%) of the main feature characterized by the first four-order EOFs in Area B is relatively lower compared with Areas A (98.98%) and C (97.12%). In addition, the SOFAR depth of the SSP within the cold eddy (cold front zone) is lower, while the SOFAR depth of the SSP within the warm eddy (warm front zone) is deeper. Influenced by the two aspects, the reconstructed SSP shows the maximum sound speed error ($\sim 5 \text{ m/s}$) around the SOFAR axis, with a SOFAR depth error of ~100 m (Figure 8b) and the RMSE of 0.98 m/s, confirming the RMSE range of Area B shown in Table 3. Then, the reconstruction error is further magnified in the sound field, which shows an obvious difference in the CZ. The TL results with SD of 300 m and frequency of 100 Hz under the background of the Argo SSP and reconstructed SSP are illustrated in Figure 9d,e. It can be found that there is a stable sound channel at a depth of 300 m, and the TL increases slightly with the increase in propagation distance as shown in Figure 9d,f. However, this stable sound channel disappears and is replaced by another sound channel with a CZ of about 40 km interval in Figure 9e,f. The comparison of the two TL curves in the RD of 300 m shows an obvious difference, that is, the red TL decreases drastically at the depths ranging from 9 km to 10 km because of the acoustic shadow zone in Figure 9e, and the CZs at the depths of about 40 km and 80 km are clearly visible.

In Area C, the difference is concentrated in the ML. The TL curves at the SD of 15 m and signal frequency of 100 Hz under the background of the Argo SSP and reconstructed SSP are illustrated in Figure 9g,h. Due to the positive gradient distribution of the Argo SSP in the ML and the big difference in gradients between the ML and the thermocline, the sound rays launched from the sound source generate the surface channel and the deep-sea channel with the CZ of ~17 km interval as shown in Figure 9g. However, owing to the reduced differences of gradients between the ML and the thermocline in the reconstructed SSP, there is only one deep-sea channel with the CZ of about 17 km interval in Figure 9h. Hence, the blue TL caused together by the surface and deep-sea channels is lower than the red TL caused solely by the deep-sea channel as shown in Figure 9i. The significant decrease in the red TL at a distance of ~4 km corresponds to the acoustic shadow zone at a depth of 30 m and distance of ~4 km in Figure 9h.



Figure 9. The sound fields with the Argo SSP (left column) and reconstructed SSP (middle column) in Areas (**a**,**b**) A, (**d**,**e**) B, and (**g**,**h**) C, with the SD of 20 m, 300 m, and 15 m, respectively. The TL comparison curves in Areas (**c**) A, (**f**) B, and (**i**) C, with the RD of 50 m, 300 m, and 50 m, respectively.

5. Conclusions

In this study, the sEOF-R method was employed for reconstructing the SSPs by joining the SLA, SSTA, and the first four-order EOFs of the historical SSPs. Three typical areas (the equator, KE, and Northeast Pacific) were selected to study the feasibility of the sEOF-R method in the SSP reconstruction. To comprehensively evaluate the performance, the RMSE and associated sound field were assessed. Among the three typical sea areas, the best performance of the sEOF-R method is obtained in the Northeast Pacific, followed by the equator area, and the worst performance occurs in the KE. In view of the sound propagation, the reconstructed SSP in the equator area is capable of reverting two sound channels of the Argo SSP. In the KE, the difference in the SOFAR depth between the Argo SSP and the reconstructed SSP is about 100 m, resulting in the disappearance of the sound channel at the depth of 300 m. In the Northeast Pacific, the surface channel disappears due to the big difference in the ML between the Argo SSP and the reconstructed SSP. In summary, the performance of the sEOF-R method in SSP reconstruction is the best in the equator region, where the hydrological structure is relatively stable. However, in areas with fruitful dynamic systems, such as the KE, the reconstruction effects are limited.

The SSP reconstruction performance results in this study demonstrate that the sEOF-R method is limited and even fails in complex and dynamic environments, such as western boundary currents, the Antarctic Circumpolar Current, and areas where cold and warm currents converge. There are two possible ways—also the future work and goals to improve the performance. One is to introduce the underwater historical parameters (e.g., thermocline) in regression analysis to improve the reconstruction accuracy [27]; the other is to establish parametric models for specific oceanic processes (e.g., mesoscale eddy and front) [10].

Author Contributions: Conceptualization, W.C. and Y.Z.; methodology, W.C.; software, W.C.; validation, Y.Z., Y.C. and S.C.; formal analysis, W.C.; investigation, W.C.; resources, W.C.; data curation, W.C.; writing—original draft preparation, W.C. and S.C.; writing—review and editing, W.C. and Y.Z.; visualization, Y.L.; supervision, L.M.; project administration, Y.C.; funding acquisition, K.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Science and Technology Innovation Program of Hunan Province (2022RC3070).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors appreciate the use of the World Ocean Atlas 2018 (WOA18) from the National Oceanic and Atmospheric Administration (NOAA) for the climatological T/S profiles (https://www.ncei.noaa.gov/archive/accession/NCEI-WOA18, accessed on 11 January 2023), the Global Ocean Argo Grid Dataset 2018 (BOA_Argo 2018) from China Argo Real-time Data Center for the gridded T/S profiles (http://www.argo.org.cn/, accessed on 11 January 2023), and Marine Environment Monitoring Service (CMEMS) for monthly average SLA and SST dataset (https://marine.copernicus.eu/, accessed on 11 January 2023).

Acknowledgments: We thank Qilin Hong for his reasonable suggestions in writing. We also thank the anonymous reviewers for their valuable suggestions to improve this work.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- 1. Brock, H.K.; Buchal, R.N.; Spofford, C.W. Modifying the sound-speed profile to improve the accuracy of the parabolic-equation technique. *J. Acoust. Soc. Am.* **1977**, *62*, 543–552. [CrossRef]
- Buckingham, M.J. Sound speed and void fraction profiles in the sea surface bubble layer. *Appl. Acoust.* 1997, 51, 225–250. [CrossRef]
- Xu, W.; Schmidt, H. System-orthogonal functions for sound velocity profile perturbation. J. Acoust. Soc. Am. 2002, 112, 2392. [CrossRef]
- 4. Vagle, S.; Burch, H. Acoustic measurements of the sound-speed profile in the bubbly wake formed by a small motor boat. *J. Acoust. Soc. Am.* **2005**, *117*, 153–163. [CrossRef]
- 5. Heimann, D.; Bakermans, M.; Defrance, J.; Kühner, D. Vertical Sound Speed Profiles Determined from Meteorological Measurements Near the Ground. *Acta Acust. United Acust.* 2007, 93, 228–240.
- Chen, C.; Ma, Y.; Liu, Y. Reconstructing Sound speed profiles worldwide with Sea surface data. *Appl. Ocean Res.* 2018, 77, 26–33. [CrossRef]
- Miyake, M.; Emery, W.J.; Lovett, J. An Evaluation of Expendable Salinity-Temperature Profilers in the Eastern North Pacific. J. Phys. Oceanogr. 1981, 11, 1159–1165. [CrossRef]
- 8. Fuda, J.L.; Millot, C.; Taupier-Letage, I.; Send, U.; Bocognano, J.M. XBT monitoring of a meridian section across the western Mediterranean Sea. *Deep Sea Res. Part I* 2000, 47, 2191–2218. [CrossRef]
- Candy, J.V.; Sullivan, E.J. Sound velocity profile estimation: A system theoretic approach. *IEEE J. Ocean. Eng.* 1993, 18, 240–252. [CrossRef]
- Chen, W.; Zhang, Y.; Liu, Y.; Ma, L.; Wang, H.; Ren, K.; Chen, S. Parametric Model for Eddies-Induced Sound Speed Anomaly in Five Active Mesoscale Eddy Regions. J. Geophys. Res. Oceans 2022, 127, e2022JC018408. [CrossRef]
- 11. Zhang, Y.; Chen, X.; Dong, C. Anatomy of a Cyclonic Eddy in the Kuroshio Extension Based on High-Resolution Observations. *Atmosphere* **2019**, *10*, 553. [CrossRef]
- 12. Chen, W.; Zhang, Y.; Liu, Y.; Wu, Y.; Zhang, Y.; Ren, K. Observation of a mesoscale warm eddy impacts acoustic propagation in the slope of the South China Sea. *Front. Mar. Sci.* 2022, *9*, 1086799. [CrossRef]

- Ivanov, E.; Capet, A.; Barth, A.; Delhez, E.J.M.; Soetaert, K.; Grégoire, M. Hydrodynamic variability in the Southern Bight of the North Sea in response to typical atmospheric and tidal regimes. Benefit of using a high resolution model. *Ocean Modell.* 2020, 154, 101682. [CrossRef]
- 14. Gettelman, A.; Geer, A.J.; Forbes, R.M.; Carmichael, G.R.; Feingold, G.; Posselt, D.J.; Stephens, G.L.; van den Heever, S.C.; Varble, A.C.; Zuidema, P. The future of Earth system prediction: Advances in model-data fusion. *Sci. Adv.* **2022**, *8*, 1–12. [CrossRef]
- Dong, R.; Leng, H.; Zhao, J.; Song, J.; Liang, S. A Framework for Four-Dimensional Variational Data Assimilation Based on Machine Learning. *Entropy* 2022, 24, 264. [CrossRef] [PubMed]
- 16. Carrassi, A.; Bocquet, M.; Bertino, L.; Evensen, G. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Clim. Chang.* **2018**, *9*, e535. [CrossRef]
- 17. Lee, Y.; Majda, A.J. State estimation and prediction using clustered particle filters. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 14609–14614. [CrossRef] [PubMed]
- 18. Li, H.; Xu, F.; Zhou, W.; Wang, D.; Wright, J.S.; Liu, Z.; Lin, Y. Development of a global gridded Argo data set with Barnes successive corrections. *J. Geophys. Res. Ocean.* **2017**, 122, 866–889. [CrossRef]
- Fox, D.N.; Teague, W.J.; Barron, C.N.; Carnes, M.R.; Lee, C.M. The Modular Ocean Data Assimilation System (MODAS). J. Atmos. Ocean. Technol. 2002, 19, 240–252. [CrossRef]
- 20. Helber, R.W.; Townsend, T.L.; Barron, C.N.; Dastugue, J.M.; Carnes, M.R. Validation Test Report for the Improved Synthetic Ocean Profile (ISOP) System, Part I: Synthetic Profile Methods and Algorithm; Naval Research Lab.: Hancock County, MS, USA, 2013.
- Park, J.C.; Kennedy, R.M. Remote sensing of ocean sound speed profiles by a perceptron neural network. *IEEE J. Ocean. Eng.* 1996, 21, 216–224. [CrossRef]
- 22. Carnes, M.R.; Teague, W.J.; Mitchell, J.L. Inference of Subsurface Thermohaline Structure from Fields Measurable by Satellite. *J. Atmos. Ocean. Technol.* **1994**, *11*, 551–566. [CrossRef]
- 23. Yan, H.Q.; Zhang, R.; Wang, H.Z.; Bao, S.L.; Chen, J.; Wang, G.J. A Surface Quasi-Geostrophic-Based Dynamical-Statistical Framework to Retrieve Interior Temperature/Salinity from Ocean Surface. *J. Geophys. Res. Ocean.* **2021**, *126*, 10. [CrossRef]
- Li, J.; Shi, Y.; Yang, Y.; Chen, C. Comprehensive Study of Inversion Methods for Sound Speed Profiles in the South China Sea. J. Ocean Univ. China 2022, 21, 1487–1494. [CrossRef]
- Lu, S.; Liu, Z.; Li, H.; Li, Z.; Wu, X.; Sun, C.; Xu, J. Manual of Global Ocean Argo Gridded Data Set (BOA_Argo) (Version 2019); China Argo Real-Time Data Center: Hangzhou, China, 2020; 14p.
- 26. Porter, M.B. The BELLHOP Manual and User's Guide PRELIMINARY DRAFT. 2011. Available online: http://oalib.hlsresearch. com/Rays/HLS-2010-1.pdf (accessed on 9 January 2023).
- Liu, C.; Han, K.; Zhang, W.; Chen, W. An Optimization Method for Sound Speed Profile Inversion Using Empirical Orthogonal Function Analysis. In Proceedings of the 2020 IEEE 3rd International Conference on Information Communication and Signal Processing (ICICSP), Shanghai, China, 12–15 September 2020.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.