*Technical Note*

# Enhancing the Assimilation of SWOT Simulated Observations Using a Multi-Scale 4DVAR Method in Regional Ocean Modeling System

Chaojie Zhou [1,2,*], Wei Cui [3], Ruili Sun [1], Ying Huang [1] and Zhanpeng Zhuang [3]

1  Hainan Institute of Zhejiang University, Sanya 572024, China; srl@zju.edu.cn (R.S.); 22234024@zju.edu.cn (Y.H.)
2  Key Laboratory of Marine Science and Numerical Modeling, Ministry of Natural Resources, Qingdao 266061, China
3  First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China; cuiwei@fio.org.cn (W.C.); zhuangzp@fio.org.cn (Z.Z.)
*  Correspondence: cjzhou@zju.edu.cn

**Abstract:** This paper presents an innovative approach to enhance the assimilation of high-resolution simulated observations, specifically targeting Surface Water Ocean Topography (SWOT) Ka-band Radar Interferometer Sea Surface Height (SSH) products, within the Regional Ocean Modeling System (ROMS). Responding to the demand for improved assimilation techniques, we developed a multi-scale Four-Dimensional Variational Data Assimilation (4DVAR) system, building upon validated fine-scale correction capabilities from prior studies. The multi-scale strategy was extended to the ROMS-4DVAR system, providing a comprehensive solution for assimilating high-resolution observations. Leveraging the Observing System Simulation Experiment (OSSE) framework, we conducted a twin experiment comprising a nature run and a free run case. Subsequently, synthetic SWOT SSH measurements were decomposed, considering the model configuration resolution. These components, derived from dense SSH observations, were integrated into a two-step 4DVAR assimilation scheme. The first cycle targets large-scale features for model field correction, and the updated analysis serves as the background for the second assimilation step, addressing fine-scale observation components. Comparisons with the primitive ROMS-4DVAR using a single-scale scheme highlight the superiority of the multi-scale strategy in reducing gaps between the model and the SSH observations. The Root Mean Squared Error (RMSE) is halved, and the Mean Absolute Percentage Error (MAPE) decreases from 2.237% to 0.93%. The two-step assimilation procedure ensures comprehensive multi-scale updates in the SSH field simulation, enhancing fine-scale features in the analysis fields. The quantification of three-dimensional-model dynamic fields further validates the efficiency and superiority of the multi-scale 4DVAR approach, offering a robust methodology for assimilating high-resolution observations within the ROMS.

**Keywords:** SWOT SSH observations; ROMS-4DVAR; multi-scale assimilation

## 1. Introduction

Numerous studies have documented the critical role of sub-mesoscale progress (~10 km) in the ocean's energy cascade [1–4]. However, existing observing approaches struggle to accurately represent the occurrence and evolution of the target continuously and accurately due to their sparse and discrete measurements. This situation would be alleviated by the launch of the Surface Water Ocean Topography (SWOT) mission [5]. The equipped Ka-band Radar Interferometer (KaRIn) measures the global Sea Surface Height (SSH) at a 21-day cycle, providing the along- and across-track SSH observations at very high resolutions (~2 km), and the Low Rate (LR) products that focus on the ocean were released in December 2023. The new-type SSH observations have been considered an

effective way to resolve signals with wavelengths down to 15 km [6,7], and they would enhance the research on sub-mesoscale processes. Recent studies have been performed to quantify the potential contribution of the SWOT observations in ocean state estimation and forecasting [8–11].

Multi-scale features, encompassing large-scale currents, mesoscale eddies, and sub-mesoscale processes, coexist in the SWOT wide-swath maps. However, the amalgamation of these diverse features may surpass the cognitive capacity of the model. Utilizing assimilation with a singular decorrelation scale, as highlighted by D'addezio et al. [12], is likely to result in information loss across other scales and may impede the model's ability to capture crucial interactions and feedback between different scales, consequently affecting the overall model performance. To address the issue of observation aliasing in data assimilation, multi-scale methodologies have been proposed and developed in previous studies [13–20]. The primary concept of this methodology involves decomposing information from high-resolution observations into large-scale and small-scale components. These components are then integrated separately with the model results at the corresponding scales or assimilated into the model through a sequential recursive process. In light of the distinct features exhibited by contemporary high-resolution observational data and the demands of assimilation, multi-scale assimilation methods are continuously refined and widely applied across diverse models. For instance, when confronted with locally dense Glider observations, Carrier et al. [17] implemented a two-step assimilation strategy within the Four-Dimensional Variational Data Assimilation (4DVAR) framework, a methodology later refined and extended to assimilate SWOT observations in the Three-Dimensional Data Assimilation (3DVAR) system [18]. Another illustrative example involves Li et al. [21], who decomposed SWOT observational data into three scale components and integrated them individually with the model results at the corresponding scales using an extended 3DVAR method.

To enhance the application of SWOT observations and advance the application of multi-scale assimilation strategy, we have implemented an extension of multi-scale assimilation methods within the ROMS-4DVAR framework, and the noteworthy positive impact of the multi-scale assimilation method is verified through an Observing System Simulation (OSSE). The paper is structured as follows: Section 2 provides a comprehensive overview of the methods employed in this study, covering the model setup, SWOT synthetic data simulation, and the implementation of the multi-4DVAR method. Section 3 presents the research findings, emphasizing assimilation method performance validation and comparisons with nature run experiments within the Observing System Simulation Experiment (OSSE) framework. Finally, Section 4 concludes the study and outlines future perspectives.

## 2. Methods

In this section, the construction details of an OSSE framework are illustrated. The primary objective was to enhance the assimilation of SWOT SSH observations. To this end, a comprehensive twin experiment, encompassing a nature run, a free run, and multiple assimilation cases, was executed as delineated below. Firstly, a nature run case constructed using a higher horizontal and vertical resolution was instantiated, and its outcomes were posited as a representation of the "true" state of the ocean. In the second step, synthetic SWOT observations were meticulously simulated by sampling the surface elevation field of the nature run at the predetermined orbit locations. Subsequently, these synthetic observations were assimilated into the free run, utilizing the primary single-scale and the proposed multi-scale ROMS-4DVAR system. Finally, the third step involved a rigorous evaluation of the impact of SWOT observations on the simulation of the ocean state.

### 2.1. Model Configuration

The Regional Ocean Modeling System (ROMS) was employed to generate three-dimensional oceanic fields and quantify the assimilation of SWOT observations within the OSSE system. The primitive equations governing ROMS were formulated under

the hydrostatic and Boussinesq assumptions. This study is centered on the Northern South China Sea (NSCS) domain, spanning from 105.5°E to 125.7°E and from 14.5°N to 26.5°N. Within the OSSE framework, a twin experiment featuring two cases was conducted, with horizontal resolution and vertical S-coordinate configuration set at $1/36° \times 1/36°$ ($1/10° \times 1/10°$) and 50 (30), respectively. The topography fields for the two cases were derived from the General Bathymetric Chart of the Oceans (GEBCO) data and the ETOPO1 global relief model integrated topography. The modeling minimum and maximum depths across the entire domain were set at 10 and 5000 m, respectively. Figure 1 illustrates the smoothed bathymetry used in the nature run. Model inputs, encompassing the initial state, surface air–sea forcing fields, and boundary conditions for both the nature run and free run, are detailed in Table 1.
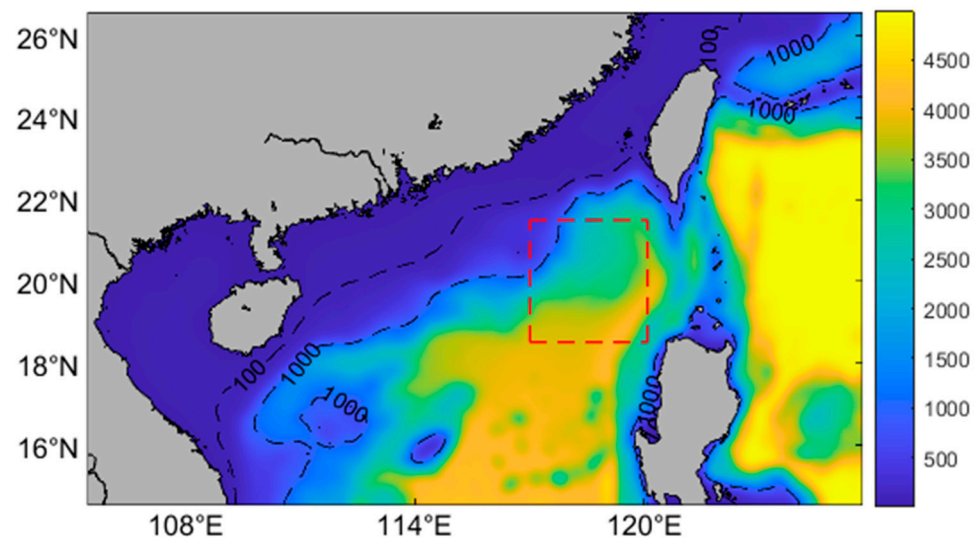


**Figure 1.** Smoothed bathymetry of nature run. The dashed contour lines indicate the 100 m and 1000 m isobaths. The dashed red box denotes the square subregion over which the power spectral density was analyzed.

**Table 1.** Configuration of model inputs for free run and nature run.

|  | Surface Forcing | Boundary | Tide | Bathymetry |
|---|---|---|---|---|
| Free Run | ECMWF (0.25°, 6 h) | SODA (0.25°, day) | No | ETOPO1 |
| Nature Run | ECMWF (0.125°, 3 h) | HYCOM (0.08°, 3 h) | OTIS | GEBCO |

The initialization of surface water elevation and three-dimensional temperature, salinity, and velocity fields for both the nature run and free run was accomplished utilizing the HYbrid Coordinate Ocean Model and Navy Coupled Ocean Data Assimilation (HYCOM and NCODA) global reanalysis dataset. For the nature run, the atmospheric forcing fields were derived from the European Centre for Medium Range Weather Forecasts (ECMWF)'s ERA-interim reanalysis [22] with a 3 h temporal resolution and a spatial resolution of $0.125° \times 0.125°$ while the free run utilized data with a 6 h temporal resolution and a spatial resolution of $0.25° \times 0.25°$. The analysis and forecast products of ERA-interim were integrated for this purpose. Boundary conditions were established using the 5-day and monthly Simple Ocean Data Assimilation (SODA) reanalysis dataset, providing temperature, salinity, sea surface height, and velocity fields. Additionally, tide forcing along the open boundaries for the nature run was incorporated from the global Oregon Tidal Inverse Solution (OTIS). The vertical mixing process was parameterized using the level 2.5 Mellor–Yamada method [23].

### 2.2. Simulated SWOT Observations

The SWOT observations were synthetically simulated using the results from the nature run. Previous investigations emphasizing SWOT Calibration/Validation (Cal/Val) underscored the critical importance of reproducing realistic internal tide energy in the nature run [8,12]. To ensure the reliability of the simulated SWOT observations, an initial exploration that focused on the SSH accuracy in the nature run case was conducted. Considering the introduction of tide forcing in the modeling process, the barotropic tidal signals involved in the surface evaluation field were removed, employing the harmonic analysis method based on the Least Squares Fit technology.
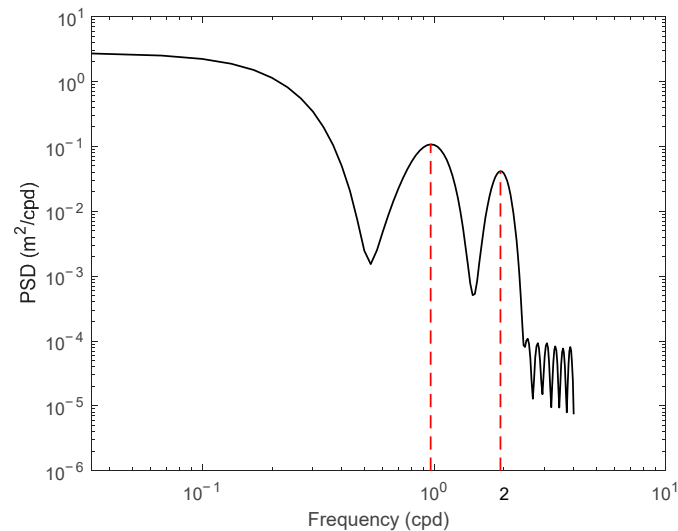
Qualitatively, the monthly mean nature SSH filed for June 2012 exhibits favorable comparability with assimilative HYCOM reanalysis data. As depicted in Figure 2a,b, the spatial patterns, including height magnitude and location, demonstrate noteworthy similarity, particularly along the Kuroshio western boundary current. The assimilation of multi-platform data such as satellite altimeter observations, in-situ temperature and salinity profiles from XBTs, Argo floats, and moored buoys by HYCOM contributes to a more realistic reconstruction of mesoscale characteristics. Additionally, we also explored the daily variation in sea surface elevation from 1 January to 31 June 2012. The spatial averaged value of the daily SSH derived by HYCOM and nature run is presented in Figure 2c. The consistent tendency in the seasonal cycle captured by the two SSH data series is observed, and the difference with the HYCOM reanalysis product had been effectively controlled since the middle of February.
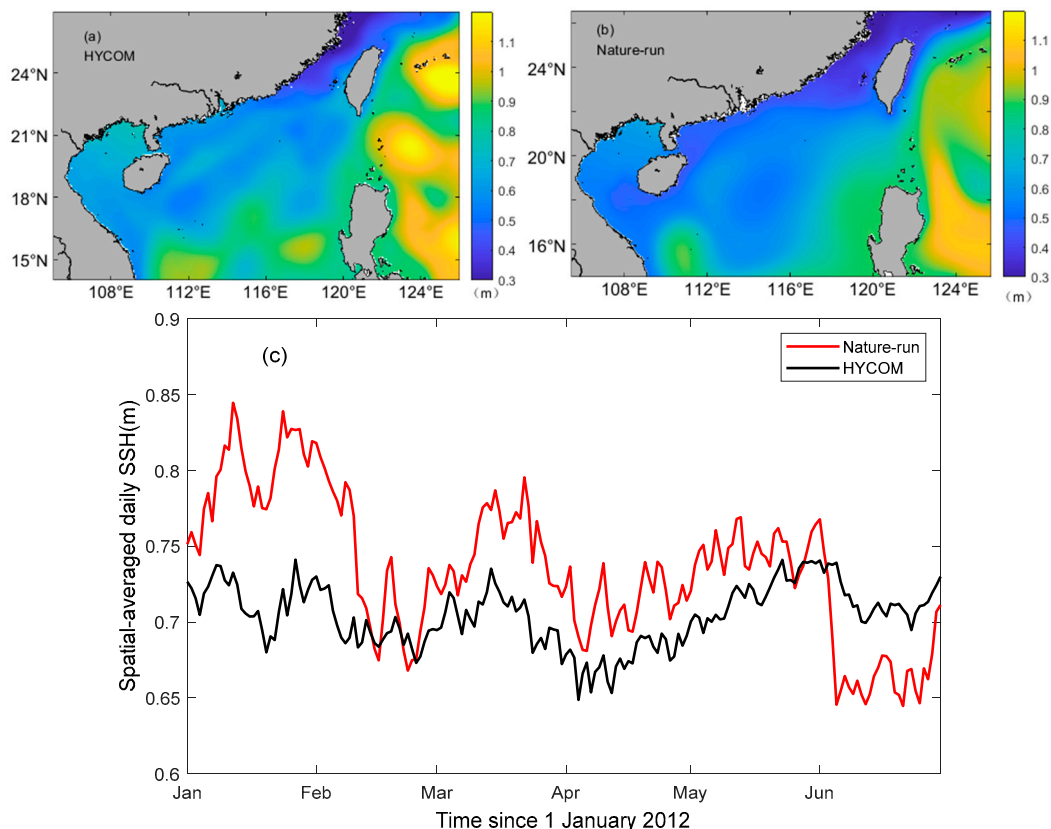


**Figure 2.** Comparison of sea surface elevation between nature run outputs and HYCOM reanalysis product. Here, (**a**,**b**) are the monthly mean SSH fields of June 2012 and (**c**) presents the temporal variation in daily SSH (after spatial variation averaged over the whole domain).

Furthermore, the internal tidal energy contained in the surface elevation field was examined. For simplicity, a subdomain proximal to the Luzon Strait, demarcated by the red box in Figure 1, was selected for analysis. The two-dimensional power spectral density

(PSD) was computed for each grid spanning the period from 1 January to 31 August 2012. The spatially averaged energy spectrum is depicted in Figure 3. The observed spectrum reveals prominent peaks near semidiurnal and diurnal frequencies. The power spectral densities adjacent to these frequencies reach values as high as $10^{-1}$ m$^2$/cpd, underscoring the substantial reproduction of primary internal tidal energy within the nature run. Consequently, the demonstrated feasibility of utilizing the SSH fields from the nature run for simulating SWOT measurements is established.



**Figure 3.** Spatial averaged energy spectrum analysis of the Luzon Strait subdomain. The Power Spectral Densities (PSDs) are calculated and illustrated in the frequency of cycle per day (cpd). The black solid line represents the analyzed PSDs, while the dashed red lines denote the 0.97 and 1.93 cpd. Two large peaks near the semidiurnal and diurnal frequencies are accurately reproduced, and there exists some spurious and weak energy at frequencies higher than 2 cpd, which is caused by the simulation errors of the nature model.

To generate the SWOT observations, we employed the SWOT-simulator software of version 1.2.4 [5] developed by the Jet Propulsion Laboratory (JPL). The simulated SWOT measurements were sampled instantaneously, utilizing the 3-hourly surface elevation outputs from the nature run. The wide-swath capability of SWOT was set at 120 km while the along- and across-track resolutions were configured typically as 2 km, consistent with previous studies [18]. Figure 4 presents the 5-day simulated SWOT observations, with detailed magnifications of measurement particulars along the wide-swath track near the Luzon Strait provided in the right panel.
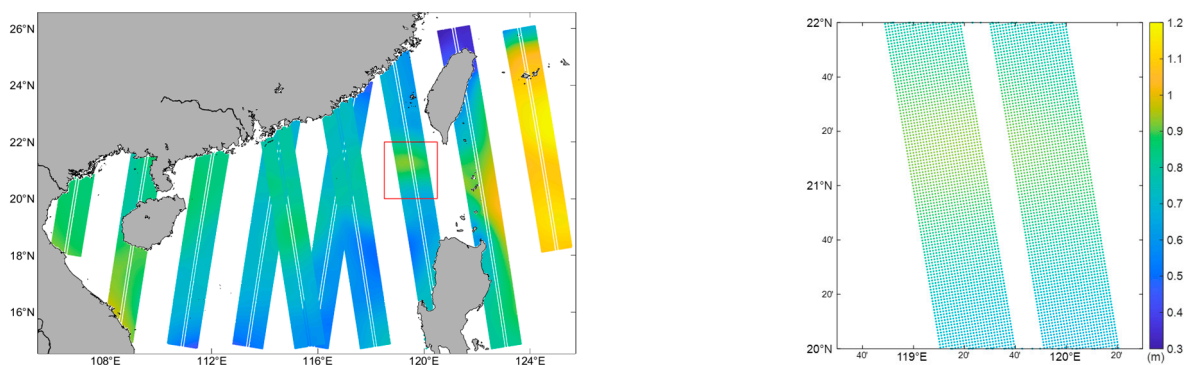


**Figure 4.** Simulated SWOT observations during the period from 5 to 10 January 2012. The enclosed box denotes the location of the magnified tracks in the right panel.

### 2.3. Data Assimilation Method

The multi-scale assimilation method used in this study was developed based on the ROMS and 4DVAR system [24]. The target of primitive 4DVAR is to identify the best estimate circulation by minimizing the difference between the model background and the observations, subject to prior hypotheses about errors and possibly additional constraints. Generally, the variational cost function for the incremental approach is given by the equation below:

$$\mathbf{J}(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^{\mathbf{T}}\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{d})^{\mathbf{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}), \tag{1}$$

Here, $\mathbf{y}$ is the observation sequence and $\mathbf{x}^b$ represents the model background state. $\mathbf{H}$ denotes the observation operator that maps the model state fields to the observation position. $\delta\mathbf{x}$ represents the increment of the state variable, surface forcing, and open boundary conditions and $\mathbf{d} = \mathbf{y} - \mathbf{H}\mathbf{x}^b$ is the innovation vector representing the difference between observations and model value. As stated by Moore et al. [24], the innovation vector $\mathbf{d} = (\cdots \mathbf{d}_i \cdots)$ represents the combination of all the observations located in the specified window, and the matrix $\mathbf{B}$ and $\mathbf{R}$ vectors are block diagonals with blocks $\mathbf{B}_i$ and $\mathbf{R}_i$, respectively.

The analysis increment $\delta\mathbf{x}^a$ derived by minimizing $\mathbf{J}(\delta\mathbf{x})$ in a least-squares sense is added to the background $\mathbf{x}^b$ to form the analysis solution $\mathbf{x}^a$, so we have

$$\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x}^a, \tag{2}$$

and the increment $\delta\mathbf{x}^a$ could be expressed as

$$\delta\mathbf{x}^{\mathbf{a}} = \mathbf{K}\mathbf{d} = \mathbf{B}\mathbf{H}^{\mathbf{T}}\left(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathbf{T}} + \mathbf{R}\right)^{-\mathbf{1}}\mathbf{d}. \tag{3}$$

Here, $\mathbf{K}$ denotes the gain matrix. According to Weaver and Courtier [25] and Moore et al. [24], the background error of ROMS-4DVAR could be expanded as follows: $\mathbf{B} = \mathbf{K}_b\mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}^{\mathbf{T}}\mathbf{K}_b^{\mathbf{T}}$. $\mathbf{K}_b$ is the multivariate balance operator, utilizing hydrostatic and geostrophic balance to constrain different model variables. In the case of rough topography, the balance operator can be disabled. Moreover, the tangent linear model and adjoint model enable the transfer of information between variables through model dynamics, and thus, disabling the usage of $\mathbf{K}_b$ would not destroy the solutions. $\mathbf{\Sigma}$ denotes the standard deviation matrix and is calculated using the free run model results for a period that is long enough to compute meaningful circulation statistics like mean and standard deviations for all prognostic state variables.

The ROMS-4DVAR system has been efficiently applied to assimilate multi-platform oceanic observations. Particularly, dense datasets originating from instruments such as gliders and the SWOT mission have significantly contributed to enhancing the correction of the model state at higher resolutions. The process of "thinning" observations has been employed as an immediate step to accommodate high-frequency sampling measurements. This involves a reduction in the number of observations within a specified radius based on the local decorrelation length scale, thereby constraining the spatial density of retained observations. Consequently, the full utility of dense measurements becomes challenging. To figure out the reasonable assimilation of observations beyond mesoscale, the reliable way is to adopt the multi-step strategy, which involves performing the large-scale corrections to the background field in the first cycle and updating the modified analysis using small-scale innovation. Several analogous multi-scale assimilation frameworks have been implemented in prior studies [8,15–21]. In this study, we developed a multi-step 4DVAR system based on ROMS, which splits the high-resolution observations into different scales and corrects the model state accordingly, thus avoiding the scale aliasing between the model background and observation fields.

In the development of the multi-step ROMS-4DVAR framework, we followed the established two-step assimilation methodology as outlined in prior literature [13,15,16]. This approach entails the assimilation of large-scale features during the initial step, followed by the adoption of the modified analysis field as the background for the second step. The overall analysis increment, denoted as $\delta\mathbf{x}^a$, is further decomposed into two distinct components, namely $\delta\mathbf{x}_L^a$ and $\delta\mathbf{x}_H^a$. These components arise from the low- and high-resolution observation assimilation cycles, respectively. Thus, we have

$$\delta\mathbf{x}^a = \delta\mathbf{x}_L^a + \delta\mathbf{x}_H^a. \tag{4}$$

In the computational process, the estimation of the analysis increment is accomplished through Equation (3). By decomposing the high-resolution observation vector, denoted as $\mathbf{y}$, into its low- and high-frequency components (designated as $\mathbf{y}_L$ and $\mathbf{y}_H$) and segregating the large- and small-scale information inherent in the model outcomes, specifically $\mathbf{x}^b = \mathbf{x}_L^b + \mathbf{x}_H^b$, the innovation vector ($\mathbf{d}_L$, $\mathbf{d}_H$) for the two-step procedure could be reformulated as $\mathbf{d}_L = \mathbf{y}_L - \mathbf{H}\mathbf{x}_L^b$ and $\mathbf{d}_H = \mathbf{y}_H - \mathbf{H}\mathbf{x}_H^b$. Consequently, the large-scale increment component can be articulated as

$$\delta\mathbf{x}_L^a = \mathbf{K}\left(\mathbf{y}_L - \mathbf{H}\mathbf{x}_L^b\right), \tag{5}$$

and the small-scale increment is expressed as

$$\delta\mathbf{x}_H^a = \mathbf{K}\mathbf{d}_H = \mathbf{K}\left(\mathbf{y}_H - \mathbf{H}\left(\mathbf{x}^b - \mathbf{x}_L^b\right)\right) = \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^b - \left(\mathbf{y}_L - \mathbf{H}\mathbf{x}_L^b\right)\right). \tag{6}$$

Supposing the large-scale field has been corrected in the first step, the term $\mathbf{y}_L$ could be estimated using $\mathbf{H}(\mathbf{x}_L^b + \delta\mathbf{x}_L^a)$. Consequently, Equation (6) can be represented as

$$\delta\mathbf{x}_H^a = \mathbf{K}\left(\mathbf{y} - \mathbf{H}\mathbf{x}^b - \mathbf{H}\delta\mathbf{x}_L^a\right) = \mathbf{K}\left(\mathbf{y} - \left(\mathbf{H}\mathbf{x}^b + \delta\mathbf{x}_L^a\right)\right). \tag{7}$$

By combining (5) and (7), the comprehensive analysis for the multi-step ROMS-4DVAR is accomplished. Notice that the increment in the first step, guided by (5), is calculated using the large-scale component of the background field. In this study, we adopted the approach introduced by [18], which approximates the increment term in (5) as the complete one. Similarly, the first-step analysis is conducted as follows:

$$\delta\mathbf{x}_L^a = \mathbf{K}\left(\mathbf{y}_L - \mathbf{H}\mathbf{x}^b\right). \tag{8}$$

Considering the limited capability of the primary model in simulating small-scale features, the biases and uncertainties introduced by this approximation are generally manageable.

For the two-step assimilation, the high-resolution SWOT observations were initially organized into bins corresponding to model grid cells. If multiple measurements fell into the same model cell, these observations were averaged to generate the "super observations". By doing this, the coupled features were split into the spatial averaged terms and the corresponding departures. Furthermore, the error covariance **C** was configured following the methodology outlined by Carrier et al. [17]. The static background error variance was computed from the history file of the preceding model run and the decorrelation scales were set proportionally to the Rossby deformation radius. Specifically, the typical averaged Rossby deformation scale in the domain is 50 km [26] and the proportionality constants used in the two-step analysis were 1.0 and 0.5 in our study.

## 3. Results

The assimilation experiment spanned from 1 January to 30 June 2012, employing both the developed multi-scale and primitive ROMS-4DVAR formulations simultaneously. The assimilation window for the two-step analysis was set at 7 days. It is worth noting that a

shorter assimilation interval could potentially enhance small-scale analysis, particularly in the second step, a facet reserved for subsequent investigations.

The analysis fields produced by both the multi-scale and primitive ROMS-4DVAR methodologies were exported at 3 h intervals. Initial evaluations entailed assessing the bias relative to the assimilated observations. Throughout the 6-month assimilation duration, over 700,000 synthetic SWOT measurements were assimilated into the 4DVAR system. Let $\{y_i^b\}$ denote the model outputs mapped to the observations' locations. The error statistical indicators, including the coefficient of determination (R2), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), were computed as follows:

$$\mathbf{R^2} = \mathbf{1} - \frac{\sum_{i=1}^{N}\left(y_i - y_i^b\right)^2}{\sum_{i=1}^{N}\left(y_i - \overline{y_i}\right)^2} \tag{9}$$

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}\left(y_i - y_i^b\right)^2} \tag{10}$$

$$\mathbf{MAPE} = \frac{100\%}{n}\sum_{i=1}^{N}\left|\frac{y_i - y_i^b}{y_i}\right| \tag{11}$$

Here, $\mathbf{N}$ is the total number of the assimilated observations, $\{y_i\}$ represents the observations, and $\overline{y_i}$ is the averaged observation value.

As illustrated in Figure 5, the model analysis states derived from both the single-scale and multi-scale assimilation approaches are consistent with the observations. However, the statistical error comparison between the two assimilation approaches indicates that the multi-scale strategy acquires better gap reduction towards the SSH observations. The coefficient of determination between the multi-scale corrected model fields and the observations has a certain improvement, from 0.991 to 0.998, the RMSE is significantly reduced by half, and the MAPE decreases from 2.237% down to 0.93%. Moreover, the points in the multi-scale panel are more concentrated near the diagonal line, and after fitting the two scattered "model-observation" data using the polynomial basis, the approximation function of the multi-scale case moves closer to the line $\mathbf{y} = \mathbf{x}$. Therefore, the enhancement optimized by the proposed multi-4DVAR scheme is validated by analyzing the residual difference between the model state and the assimilated observations.
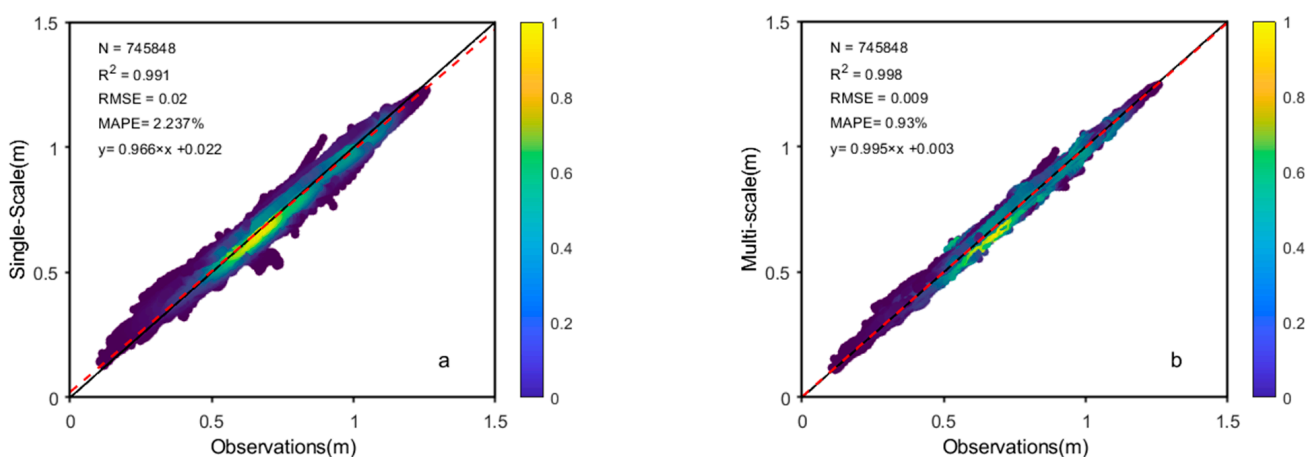


**Figure 5.** The residual difference between the observations and model state value wherein the latter had been corrected by the single-scale ROMS-4DVAR (**a**) and the proposed multi-scale 4DVAR scheme (**b**). The color represents the data density and was calculated by counting the number of scattered points located in each meshed grid, while the solid (black) and dashed (red) lines denote the diagonal line $\mathbf{y} = \mathbf{x}$ and the fitted curves of the scattered data.

Moreover, the error inter-comparison between single-scale and multi-scale ROMS-4DVAR schemes was conducted and is presented. As illustrated in Figure 6, the bias of the single-scale approach is located at [−0.1, 0.1] while the one of the multi-scale 4DVAR is reduced significantly and restricted in the interval [−0.04, 0.04], indicating that the posterior analysis after the two-step correction comes closer to the target observation than the primitive 4DVAR. Given the two-step assimilation process, wherein large-scale features are corrected in the initial round and the subsequent analysis step primarily focuses on updating small-scale features, we delve into the respective increments for the two assimilation cycles. The results, presented in Figure 7, encompass both the large-scale (a) and fine-scale (b) components of increments along with their summation (c). The two-step assimilation procedure ensures the comprehensive updating of multi-scale features in the SSH field, thereby circumventing the scale aliasing of errors mentioned in a prior study [12].
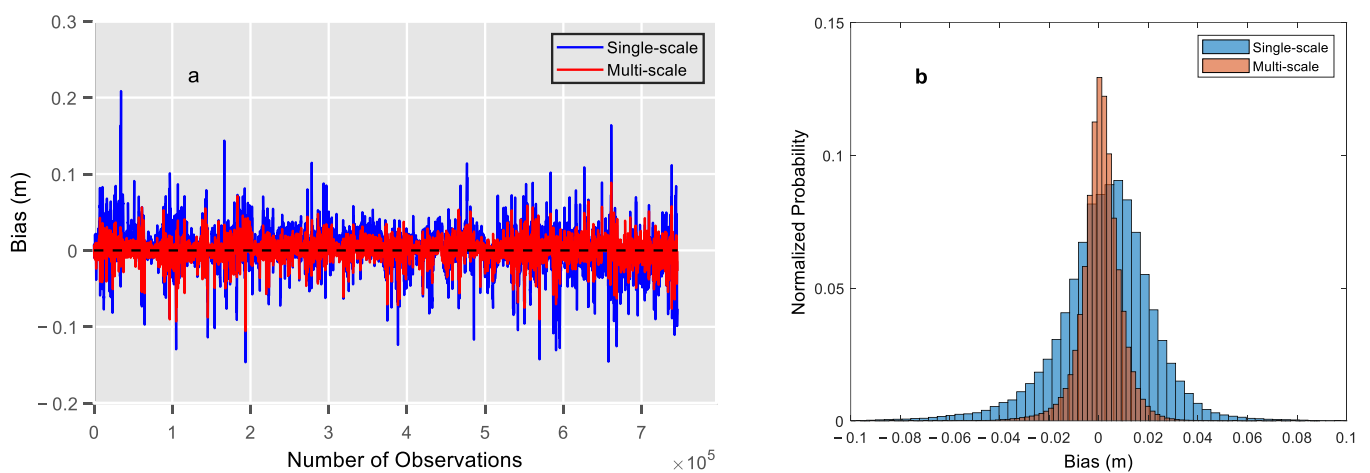


**Figure 6.** The bias comparison between single-scale and multi-scale components. The left panel (**a**) shows the residual difference of the posterior surface elevation analysis at the observation position while the right panel (**b**) compares the bias distribution of the two assimilation schemes wherein the error has been normalized.
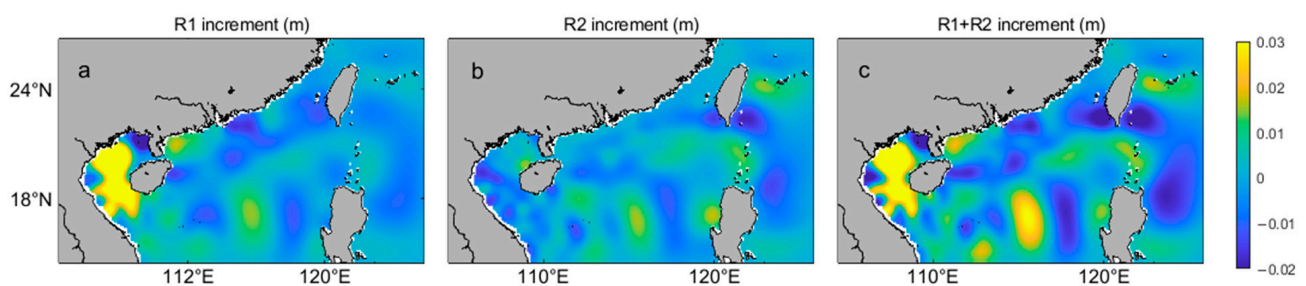


**Figure 7.** The increments of the sea surface elevation field derived from the two-step assimilation process on 24 June 2012. Here, (**a**) is the increment after the first assimilation cycle (R1), (**b**) gives the net increment of the second step (R2), and (**c**) shows the total variation introduced by the multi-scale 4DVAR assimilation.

The robustness of the proposed multi-scale modification for the ROMS-4DVAR system has been substantiated through an examination of the congruence between the corrected model state and the assimilated observations. To further assess the efficacy of the updated surface elevation fields achieved through multi-scale assimilation, the true state generated by the nature run is employed here. An analysis of the errors in surface elevation concerning the nature run from 1 January to 31 June is presented in Figure 8. Firstly, the high-resolution SSH field of the nature run series was mapped onto the coarse grid of the

free run. Subsequently, the averaged mean absolute error (MAE) was calculated across the entire region for the free-run, single-scale, and multi-scale 4DVAR assimilation scenarios. The temporal progression of the MAE underscores the beneficial impact of the assimilation procedure on model state correction, evident in both the single-scale and multi-scale strategies. The accuracy of SSH fields following the first assimilation step was comparable to that derived from the single-scale analysis and, in certain instances, exhibited a slight superiority. Furthermore, it is noteworthy that the single-scale analysis may encounter challenges in specific situations, as illustrated by the dashed rectangle in Figure 8. These arise when multi-scale features, aliased in high-resolution SWOT measurements, are directly injected into the model, leading to a subsequent negative correction. In contrast, for the multi-scale assimilation process, a global enhancement in accuracy was observed, decreasing from 3.97 cm to 3.04 cm following the correction of large-scale features. A notable error reduction of 1.0 cm was noted after the assimilation of fine-scale information.
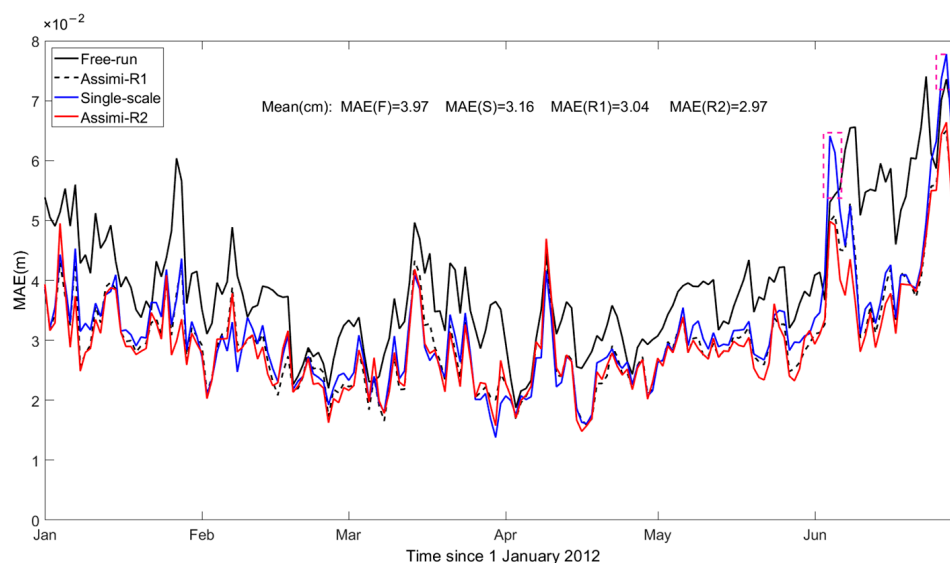


**Figure 8.** The regional averaged mean absolute error (MAE) with respect to the nature run case and the time series covering the period from January to June. The SSH accuracy values of the free run and the single-scale and multi-scale assimilation were evaluated.

Moreover, to delineate the distinctive performances of the single- and multi-scale assimilation schemes, a spatial analysis of errors was undertaken. Initially, a representative scenario depicting horizontal SSH updates is presented in Figure 9, with a comparison between the free run and nature run serving as a reference for comprehensive evaluation. In summary, the findings align with the earlier regional averaged mean absolute error (MAE) analysis in the time series, indicating overall improvements in SSH analysis fields across the entire region for all assimilation cases.

The efficacy of the single-scale method is comparable to that achieved in the first assimilation step using the multi-scale scheme. This suggests that, when adopting a single-scale assimilation, efficient absorption is confined to the large-scale features present in the observations. Furthermore, the analysis outcomes following the two-step assimilation processes demonstrate the highest accuracy, with the second assimilation cycle primarily contributing to modifications at fine scales, particularly in regions surrounding islands. As an illustrative example, the analyzed SSH fields in the vicinity of Taiwan Island exhibit a substantial reduction in the number of errors. This notable improvement is particularly noteworthy given the challenges arising from the intricate topography and the coupling of multi-scale ocean dynamic activities including tides, Kuroshio, and coastal currents. The diminished error in these areas underscores the efficiency of the second-step assimilation in updating smaller-scale patterns. Table 2 provides the RMSE and MAE for each experimental

group. The accuracy values for the first step of multi-scale and single-scale assimilation are equivalent, with MAE and RMSE values of approximately 5.3 cm and 8.3 cm, respectively.
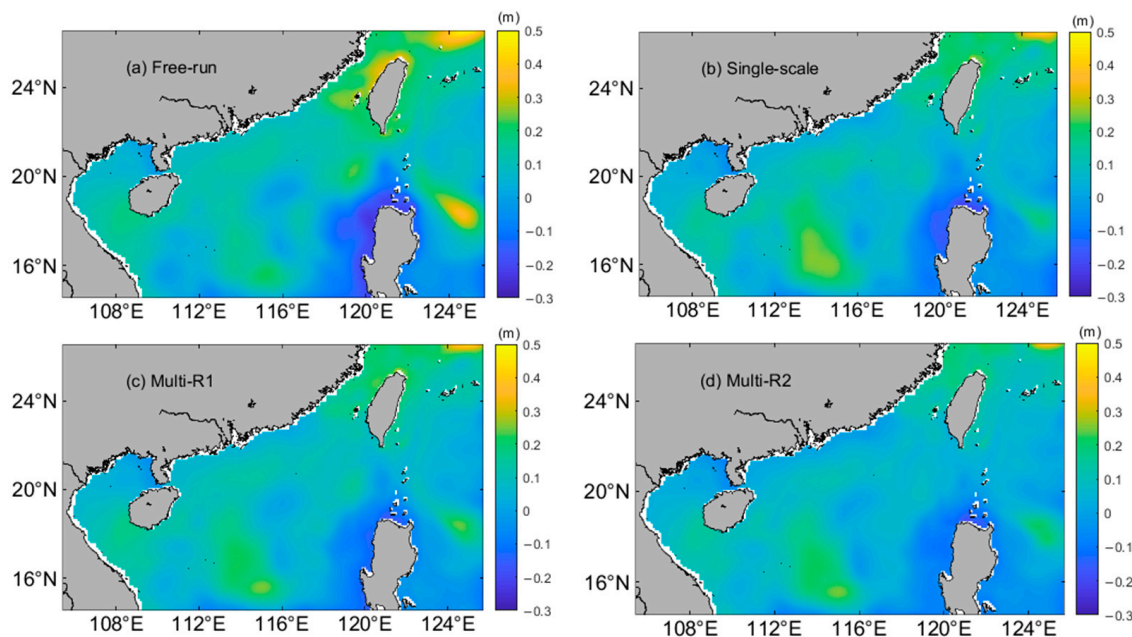


**Figure 9.** The horizontal SSH updates on 8 June after the single-scale and two-step assimilation processes. As a reference, the error result of the free run is provided.

**Table 2.** MAE and RMSE errors for each group experiment. The presented indicators were calculated using the daily averaged fields from 8 June.

| cm | Free-Run | Single-Scale | Multi-R1 | Multi-R2 |
|---|---|---|---|---|
| MAE | 6.55 | 5.25 | 5.3 | 4.35 |
| RMSE | 10.15 | 8.37 | 8.28 | 7.01 |

The comprehensive error analysis of the ocean state, encompassing SSH, three-dimensional (3D) temperature (temp), salinity (salt), and the eastern and northern velocity components (u and v) fields resulting from the assimilation cases, was systematically undertaken. To emphasize the enhancements and facilitate a comparative assessment of the impacts between single-scale and multi-scale assimilation, the error reduction was quantified relative to the accuracy of the free run case. The MAE and RMSE reductions are presented in Figure 10. Regarding the 3D temperature and salinity fields, the influence of SSH assimilation was negligible, as the reductions in MAE and RMSE were minimal, occasionally displaying negative values. In alignment with prior studies [27–29], the direct assimilation of altimetry measurements proved insufficient for enhancing the estimation of temperature and salinity. Additional processes are typically required before SSH data assimilation, although this aspect was not the primary focus of the present study and was left for future exploration. In addition to the evaluation of water properties, the dynamic model state variables, including surface elevation and momentum components, demonstrated significant improvement. Remarkably, the two-step assimilation within the multi-scale process produced better results compared to the single-scale approach, reinforcing the conclusions previously mentioned.
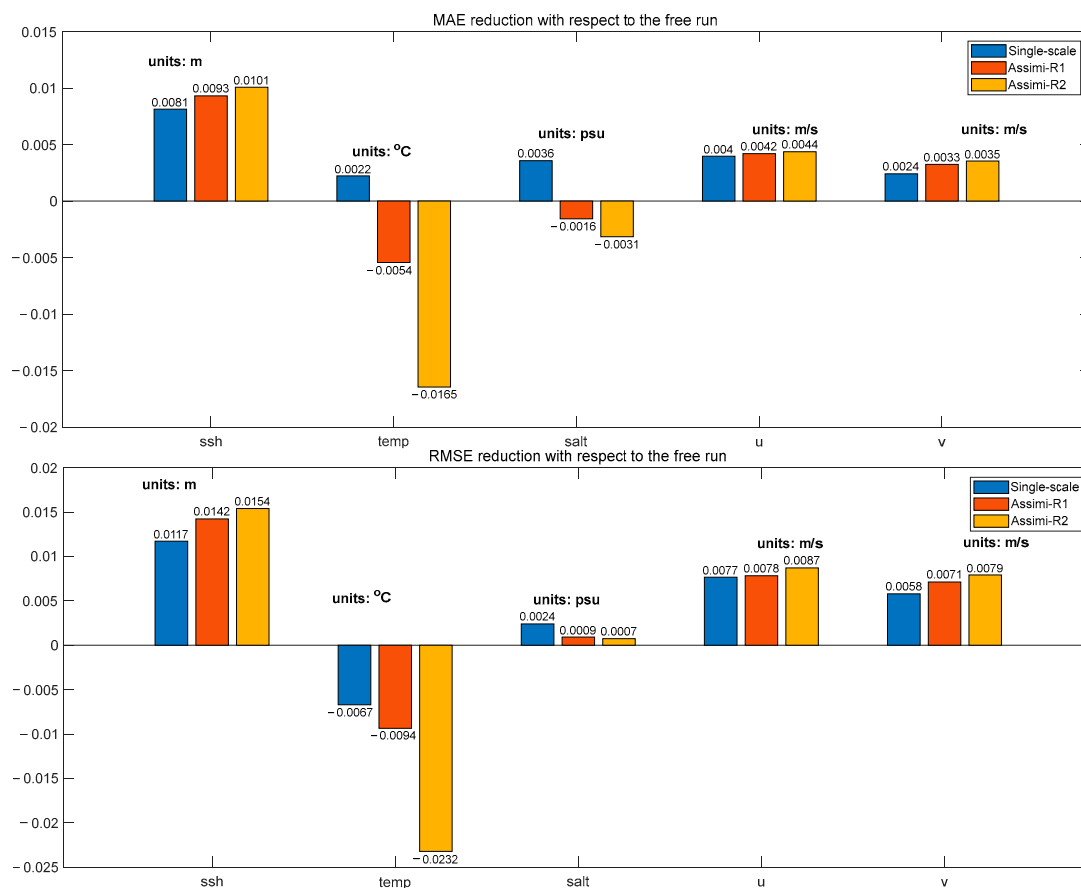
**Figure 10.** The averaged error reduction (MAE: **top**, RMSE: **bottom**) in the ocean state (SSH, temperature, salinity, eastern and northern velocity components), wherein the error decreases are analyzed with respect to the accuracy of the free run case.

## 4. Summary and Discussion

To address the imperative of assimilating high-resolution observations such as the SWOT KaRIn SSH products, this study introduced a multi-scale 4DVAR system within the ROMS framework. Building upon the validated fine-scale correcting capabilities of multi-scale assimilation schemes [15–18], we extended this approach to the ROMS-4DVAR system. Employing the OSSE framework, a twin experiment comprising a nature run and a free run case was executed. The SSH fields estimated by the nature run were validated to be comparable with HYCOM reanalysis products, affirming their reliability in simulating realistic internal tide energy and their applicability for reproducing SWOT observations.

Subsequently, synthetic SWOT SSH measurements were decomposed into spatially averaged terms and their corresponding departures, accounting for the model configuration resolution. These components, derived from dense SSH observations, were then integrated into the proposed two-step 4DVAR assimilation scheme. The first cycle involved large-scale features to correct the model fields, and the updated analysis served as the background for the second assimilation step, targeting the fine-scale observation component. A comparison with primitive ROMS-4DVAR using a single-scale scheme revealed that the multi-scale strategy achieved better gap reduction in SSH observations. The RMSE was significantly halved, and the MAPE decreased from 2.237% to 0.93%. The two-step assimilation process ensured comprehensive multi-scale updates in the SSH field, enhancing the fine-scale features in the analysis fields. A quantitative comparison with the nature run further validated the efficiency and superiority of the multi-scale 4DVAR approach, demonstrating both accuracy enhancement and fine-scale correction. Moreover, the equivalence in accuracy

between the first step of the multi-scale and single-scale approaches suggests that the single-scale 4DVAR system predominantly captures large-scale features.

The straight assimilation of SSH barely has an impact on improving the oceanic 3D temperature and salinity fields, and thus, it is essential to project the surface dynamic measurements onto the water property fields in advance. However, the strategy of dense SSH observation projection would bring other issues and needs further consideration of features such as the quality of produced 3D temperature and salinity profiles, the decomposition of the generated high-resolution underwater pseudo-observations, the necessary modification of the multi-scale 4DVAR approach, and so on. In addition, the updates of the spatial–temporal windows and the background error variance during the assimilation process should be further considered, especially for the fine-scale feature correction process in the multi-scale assimilation method. Moreover, it is also worth trying to decompose high-resolution observations into more components with different scales in order to optimize the multi-scale feature of the model results more precisely. Above all, the verification in this study was performed in the OSSE framework, and the beta products of the SWOT KaRIn L2 SSH have been available since December 2023, so the main focus in the future will be to explore the assimilation of real SWOT observation data.

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

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author, Chaojie Zhou, upon reasonable request. The data are not publicly available due to privacy restrictions.

**Conflicts of Interest:** The authors declare no conflicts 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. McWilliams, J.C. Submesoscale currents in the ocean. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *472*, 20160117. [CrossRef] [PubMed]
2. Rocha, C.B.; Gille, S.T.; Chereskin, T.K.; Menemenlis, D. Seasonality of submesoscale dynamics in the Kuroshio Extension. *Geophys. Res. Lett.* **2016**, *43*, 11–304. [CrossRef]
3. Su, Z.; Wang, J.; Klein, P.; Thompson, A.F.; Menemenlis, D. Ocean submesoscales as a key component of the global heat budget. *Nat. Commun.* **2018**, *9*, 775. [CrossRef] [PubMed]
4. Zhang, Z.; Liu, Y.; Qiu, B.; Luo, Y.; Cai, W.; Yuan, Q.; Tian, J. Submesoscale inverse energy cascade enhances Southern Ocean eddy heat transport. *Nat. Commun.* **2023**, *14*, 1335. [CrossRef] [PubMed]
5. Gaultier, L.; Ubelmann, C.; Fu, L.L. The challenge of using future SWOT data for oceanic field reconstruction. *J. Atmos. Ocean. Technol.* **2016**, *33*, 119–126. [CrossRef]
6. Ubelmann, C.; Dibarboure, G.; Dubois, P. A cross-spectral approach to measure the error budget of the SWOT altimetry mission over the Ocean. *J. Atmos. Ocean. Technol.* **2018**, *35*, 845–857. [CrossRef]
7. Wang, J.; Fu, L.L.; Torres, H.S.; Chen, S.; Qiu, B.; Menemenlis, D. On the spatial scales to be resolved by the surface water and ocean topography Ka-band radar interferometer. *J. Atmos. Ocean. Technol.* **2019**, *36*, 87–99. [CrossRef]
8. Li, Z.; Wang, J.; Fu, L.L. An observing system simulation experiment for ocean state estimation to assess the performance of the SWOT mission: Part 1—A twin experiment. *J. Geophys. Res. Ocean.* **2019**, *124*, 4838–4855. [CrossRef]

9. Benkiran, M.; Ruggiero, G.; Greiner, E.; Le Traon, P.Y.; Rémy, E.; Lellouche, J.M.; Tchonang, B. Assessing the impact of the assimilation of swot observations in a global high-resolution analysis and forecasting system part 1: Methods. *Front. Mar. Sci.* **2021**, *8*, 691955. [CrossRef]
10. Tchonang, B.C.; Benkiran, M.; Le Traon, P.Y.; Jan van Gennip, S.; Lellouche, J.M.; Ruggiero, G. Assessing the Impact of the Assimilation of SWOT Observations in a Global High-Resolution Analysis and Forecasting System—Part 2: Results. *Front. Mar. Sci.* **2021**, *8*, 687414. [CrossRef]
11. Archer, M.R.; Li, Z.; Wang, J.; Fu, L.L. Reconstructing Fine-Scale Ocean Variability via Data Assimilation of the SWOT Pre-Launch In Situ Observing System. *J. Geophys. Res. Ocean.* **2022**, *127*, e2021JC017362. [CrossRef]
12. D'addezio, J.M.; Smith, S.; Jacobs, G.A.; Helber, R.W.; Rowley, C.; Souopgui, I.; Carrier, M.J. Quantifying wavelengths constrained by simulated SWOT observations in a submesoscale resolving ocean analysis/forecasting system. *Ocean Model.* **2019**, *135*, 40–55. [CrossRef]
13. Xie, Y.; Koch, S.; McGinley, J.; Albers, S.; Bieringer, P.E.; Wolfson, M.; Chan, M. A space–time multiscale analysis system: A sequential variational analysis approach. *Mon. Weather Rev.* **2011**, *139*, 1224–1240. [CrossRef]
14. Muscarella, P.A.; Carrier, M.; Ngodock, H. Examination of a Multi-Scale Three-Dimensional Variational Data Assimilation Scheme in the Kuroshio Extension Using the Naval Coastal Ocean Model. *Cont. Shelf Res.* **2014**, *73*, 41–48. [CrossRef]
15. Li, Z.; McWilliams, J.C.; Ide, K.; Farrara, J.D. A multiscale variational data assimilation scheme: Formulation and illustration. *Mon. Weather Rev.* **2015**, *143*, 3804–3822. [CrossRef]
16. Li, Z.; McWilliams, J.C.; Ide, K.; Farrara, J.D. Coastal ocean data assimilation using a multi-scale three-dimensional variational scheme. *Ocean Dyn.* **2015**, *65*, 1001–1015. [CrossRef]
17. Carrier, M.J.; Osborne, J.J.; Ngodock, H.E.; Smith, S.R.; Souopgui, I.; D'Addezio, J.M. A multiscale approach to high-resolution ocean profile observations within a 4DVAR analysis system. *Mon. Weather Rev.* **2019**, *147*, 627–643. [CrossRef]
18. Souopgui, I.; D'Addezio, J.M.; Rowley, C.D.; Smith, S.R.; Jacobs, G.A.; Helber, R.W.; Osbome, J.J. Multi-scale assimilation of simulated SWOT observations. *Ocean Model.* **2020**, *154*, 101683. [CrossRef]
19. Chamberlain, M.A.; Oke, P.R.; Brassington, G.B.; Sandery, P.; Divakaran, P.; Fiedler, R.A.S. Multiscale data assimilation in the Bluelink ocean reanalysis (BRAN). *Ocean Model.* **2021**, *166*, 101849. [CrossRef]
20. Chamberlain, M.A.; Oke, P.R.; Fiedler, R.A.; Beggs, H.M.; Brassington, G.B.; Divakaran, P. Next generation of Bluelink ocean reanalysis with multiscale data assimilation: BRAN2020. *Earth Syst. Sci. Data* **2021**, *13*, 5663–5688. [CrossRef]
21. Li, Z.; Archer, M.; Wang, J.; Fu, L.L. Formulation and demonstration of an extended-3DVAR multi-scale data assimilation system for the SWOT altimetry era. *Ocean Sci. Discuss.* **2021**, *2021*, 1–26.
22. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Vitart, F. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 553–597. [CrossRef]
23. Mellor, G.L.; Yamada, T. Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.* **1982**, *20*, 851–875. [CrossRef]
24. Moore, A.M.; Arango, H.G.; Broquet, G.; Powell, B.S.; Weaver, A.T.; Zavala-Garay, J. The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: Part I–System overview and formulation. *Prog. Oceanogr.* **2011**, *91*, 34–49. [CrossRef]
25. Weaver, A.; Courtier, P. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.* **2001**, *127*, 1815–1846. [CrossRef]
26. Chelton, D.B.; DeSzoeke, R.A.; Schlax, M.G.; El Naggar, K.; Siwertz, N. Geographical variability of the first baroclinic Rossby radius of deformation. *J. Phys. Oceanogr.* **1998**, *28*, 433–460. [CrossRef]
27. Ezer, T.; Mellor, G.L. Continuous Assimilation of Geosat Altimeter Data into a Three-Dimensional Primitive Equation Gulf Stream Model. *J. Phys. Oceanogr.* **1994**, *24*, 832–847. [CrossRef]
28. Cooper, M.; Haines, K. Altimetric assimilation with water property conservation. *J. Geophys. Res. Ocean.* **1996**, *101*, 1059–1077. [CrossRef]
29. Alves, J.O.S.; Haines, K.; Anderson, D.L.T. Sea Level Assimilation Experiments in the Tropical Pacific. *J. Phys. Oceanogr.* **2001**, *31*, 305–323. [CrossRef]