https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

Open Access Earth System Science Data Discussions

Global submesoscale eddy identification and characteristic analysis based on multi-source remote sensing data

Meng Hou[1], Nan Luo[1], Jie Yang[1,2], Ge Chen[1,2]

[1]State Key Laboratory of Physical Oceanography, Department of Marine Technology, Ocean University of China, Qingdao, 266000, China
[2]Laboratory for Regional Oceanography and Numerical Modeling, Department of Ocean Big Data and Prediction, Laoshan Laboratory, Qingdao, 266000, China

Correspondence to: Jie Yang (yangjie2016@ouc.edu.cn) and Ge Chen (gechen@ouc.edu.cn)

Abstract. Research on oceanic submesoscale eddies has long been limited by the spatiotemporal resolution constraints of satellite altimeters. Relevant studies remain insufficient and have mostly focused on regional waters. Global investigations on submesoscale eddy detection, dataset construction, and distribution characteristic analysis based on multi-source remote sensing data and multiple methods are still lacking. In this study, we first develop a submesoscale eddy detection method by integrating high spatiotemporal resolution ocean color data, deep learning algorithms, and digital image processing techniques and construct a global submesoscale eddy dataset based on chlorophyll-a observations. Furthermore, we design a multi-scale eddy detection framework using altimeter data and establish two global datasets: a submesoscale eddy dataset derived from SWOT satellite altimeter measurements and a multi-scale eddy dataset generated from merged altimeter data. Finally, we statistically compare the scale, seasonal, and geographical distribution characteristics of eddies from the three constructed datasets and systematically analyze the strengths and limitations of different data sources and algorithms for submesoscale eddy detection. This study effectively compensates for the scarcity of existing submesoscale eddy datasets, provides a valid verification approach for conventional eddy products, and offers a feasible guideline for data selection in diverse submesoscale eddy research scenarios.

1 Introduction

Submesoscale eddies are a ubiquitous type of rotating water masses across global oceans and represent a distinctive yet pervasive form of seawater motion. Their lifespans range from several hours to a few days, with horizontal scales varying from hundreds of meters to tens of kilometers. Characterized by intense vertical motions, these eddies can penetrate to depths of over one thousand meters. Compared to mesoscale eddies, submesoscale eddies are more numerous and widely distributed, with vertical transport rates exceeding those of mesoscale eddies by an order of magnitude, enabling the upward delivery of nutrients from the deep ocean into the euphotic zone (McWilliams et al., 2016; Lévy et al., 2012; Hou et al., 2024). They play a pivotal role in disrupting ocean pycnocline, establishing vertical ecological corridors, enhancing primary

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

production, and regulating marine biogeochemical cycles. Therefore, the submesoscale eddy identification has become a critical research area in oceanography (McWilliams, 2016; Zhang et al., 2020).

Most of the existing methods for identifying mesoscale eddies can also be applied to the identification tasks of submesoscale eddies. Traditional mesoscale eddy detection methods are primarily numerical simulation algorithms based on sea surface height data. These approaches identify ocean eddies by applying fixed thresholds and discriminant criteria, and a variety of relevant algorithms have been developed over the years. Representative techniques include the Okubo-Weiss parameter method (Chelton et al., 2007), the Winding Angle method (Chaigneau et al., 2008), the Flow Direction method (Williams et al., 2011), the Sea Surface Topology method (Chelton et al., 2011b), and the Lagrangian coherent structure method (Haller et al., 2015). These methods have been evolving for decades and have yielded substantial achievements in related fields (McWilliams et al., 1990; Chelton et al., 2007; Chelton et al., 2011; Chaigneau et al., 2008; Chen et al., 2019; Xing et al., 2023). Among them, the most representative is the sea surface topology method. The sea surface topology method leverages the closed topological characteristics of eddies. It takes local extrema of sea surface height (SSH) as eddy centers and defines the outermost closed SSH contour satisfying screening criteria as eddy boundaries. It enables long-term and continuous tracking of mesoscale eddies across the global ocean, allowing researchers to capture their full life cycles and evolutionary patterns (Chelton et al., 2011). Some scholars have further optimized this approach. For instance, Mason replaced pixel-based detection with isoline interpolation and imposed strict constraints on the shape of closed contours, which effectively improved the positioning accuracy of eddy boundaries (Mason et al. 2014). In recent years, with the advancement of observation techniques and the improving spatial resolution of SSH data, eddy detection based on the sea surface topology method has attracted growing attention and become a widely adopted approach. The global mesoscale eddy dataset established by Chelton et al. using this method is now one of the most classic and frequently used references in ocean eddy research (Chen et al., 2019; Xing et al., 2023).

With the increase in computational power and deep learning technologies, some scholars have begun to explore the use of deep learning algorithms and high-resolution marine ocean color data to identify submesoscale eddies. The distinctive elliptical patches or bands formed by tracers like phytoplankton and bio-oil films on the sea surface observed in ocean color and SAR imagery serve to outline eddy configurations (Pegau et al., 2002; Liu et al., 2015; Konik et al., 2016; Chen et al., 2019; Ni et al., 2021). In addition, high-resolution radar systems in specific areas can also be utilized to detect submesoscale eddies (Liu et al., 2021). Some scholars have achieved remarkable results in the field of submesoscale eddy identification by using a deep learning method. Li proposed a multispectral eddy detection network (MED-Net), which uses multispectral data to detect oceanic submesoscale eddies and constructed a global coastal submesoscale eddy dataset (Li et al., 2024). Wang used the chlorophyll data produced by the Geostationary Ocean Color Imager (GOCI) and the Yolov7x neural network model identified abundant submesoscale eddies in the Northwest Pacific Ocean (Wang et al., 2024). Liu proposed a cross-domain neural network based on high-frequency radar (HFR) data to detect submesoscale eddies (Liu et al., 2021).

Although these methods have shown remarkable performance in detecting submesoscale ocean eddies, there are still some challenges that remain to be resolved. First, for altimetry data, the identification and study of submesoscale eddies have long

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

been limited by their spatial resolutions. In particular, research on the global-scale identification, dataset construction, and characterization of submesoscale eddies at high spatiotemporal resolution remains insufficient. Although some scholars have identified a large number of submesoscale eddies using SSH data from SWOT with a spatial resolution of 2 km, the 21-day temporal resolution of SWOT still cannot achieve daily global coverage. Consequently, relying solely on SWOT satellite data presents challenges for the detailed characterization of submesoscale eddies, including tasks such as eddy tracking and lifespan analysis. Second, for the integration of ocean color data and deep learning, existing approaches only employ rectangular bounding boxes to mark detected eddies rather than delineating their actual contours. Since oceanic eddies feature irregular non-circular shapes, rectangle-based labeling cannot fully represent their unique morphological characteristics. Therefore, it is essential to develop a method that can precisely detect global submesoscale eddies and extract their complete contours. Third, whether it is altimeter data or ocean color data, current research on submesoscale eddies mostly focuses on individual cases or local areas, while the identification, dataset establishment, and feature research of global submesoscale eddies are still insufficient (Ni et al., 2021; Wang et al., 2024; Li et al., 2024; Verger-Miralles et al., 2025). Fourth, in an era of diverse data and methods, it is difficult for a single data source to fully characterize submesoscale eddies. The performance of each data type and method varies in submesoscale eddy identification tasks, and their differences, as well as respective advantages and limitations, remain unclear and require further comparative analysis.

Against this backdrop, this paper achieves, for the first time, the establishment of three distinct global submesoscale eddy datasets based on various data sources and methodologies. The characteristics distribution of these eddies are statistically evaluated, compared, and interpreted. This effort not only compensates for the existing deficiencies in submesoscale eddy datasets but also enables readers to better comprehend the specific characteristics, strengths, and weaknesses associated with each dataset and methodological approach in the mission of submesoscale eddy identification. In this study, we defined eddies with a radius less than 25 km as submesoscale eddies. The specific work is as follows: (1) We developed a global method for detecting submesoscale eddies, which leverages the D0eepLabv3+ deep learning algorithm alongside chlorophyll data and digital image processing techniques. The findings confirm that this method is capable of reliably detecting submesoscale eddies in chlorophyll fields—ranging from several to tens of kilometers in scale—as well as delineating their complex boundaries, surpassing simple existence identification. (2) We developed an eddy identification method based on SWOT and its merged altimetry data. This method references the closed-contour approach proposed by Chelton et al. and further incorporates vorticity calculations within the closed contours as a secondary validation step, thereby enabling the effective global identification of submesoscale eddies. Using the proposed method, we constructed two separate eddy datasets: one from SWOT satellite altimetry data and the other from gridded altimetry data at a spatial resolution of 0.125°. (3) Furthermore, we conducted a comparative analysis of the global eddy datasets derived from multiple data sources and methods. We examined the distribution characteristics of submesoscale eddies identified from different datasets and elucidated the respective advantages and limitations of each dataset and method in submesoscale eddy identification tasks.

This study represents the first effort to achieve global submesoscale eddy identification, dataset construction, and comparative feature analysis using multiple data sources and methods. On the one hand, it diversifies the approaches

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

available for submesoscale eddy detection and effectively addresses the current scarcity of comprehensive submesoscale eddy datasets. On the other hand, it enhances our understanding of the distinct characteristics of different data sources and methods in submesoscale eddy identification tasks, providing new insights for researchers to select the most appropriate data for their specific studies on submesoscale eddies. Collectively, this work will further advance our understanding of submesoscale eddies.

2 Materials

2.1 Chlorophyll products

The sea surface chlorophyll data used in this study were derived from two primary sources: the Geostationary Ocean Color Imager (GOCI) and the Ocean and Land Color Instrument (OLCI) aboard the Sentinel-3A&B satellites. The high-resolution Level-2 chlorophyll data from these sensors were employed for the automatic detection of submesoscale eddies using deep learning and digital image processing techniques. The combination of two different types of data can help enhance the robustness of deep learning models. GOCI is a geostationary Earth-orbiting satellite with a viewing center located at 130°E and 36°N. It provides remote sensing imagery at a spatial resolution of 500 meters, covering an observational area of 2500 × 2500 km² (Wang et al., 2024). Owing to the adoption of a geostationary orbit, GOCI is capable of persistently observing a fixed geographical area and delivering multiple chlorophyll images each day. This capability serves as a significant asset for the sustained observation and tracking of submesoscale eddies and their related ecological processes. As part of the ESA's Copernicus program, the Sentinel-3 mission includes two satellites, Sentinel-3A and Sentinel-3B, which supply high-quality ocean color data for marine forecasting and climate research. The OLCI sensors carried by Sentinel-3A&B satellites provide data with a spatial resolution of 300 meters. And the cooperation of two satellites enables the acquisition of daily chlorophyll images for all sea areas around the world, which can be widely used to extract information about submesoscale eddies in the ocean and inland waters (Clerc et al., 2020; Sukigara et al., 2022).

We finally used three years of chlorophyll data provided by Sentinel-3A/B to identify submesoscale eddies from January 1, 2020, to December 31, 2022, and ultimately constructed a global submesoscale eddy dataset based on chlorophyll data. Furthermore, the chlorophyll data provided by GOCI was used to enhance the robustness of the neural network training outcomes. The two types of chlorophyll data utilized in this study can be accessed at https://oceandata.sci.gsfc.nasa.gov/directdataaccess/Level-2 and https://oceandata.sci.gsfc.nasa.gov/directdataaccess/Level-2/GOCI/.

2.2 Sea surface height and temperature products

As the anomalies of SSH and sea surface height SST serve as key characteristics of eddies. In this study, the SSH and SST data were used as comparative references to assess the identification results of submesoscale eddies based on chlorophyll data, and the SSH data were also used to build submesoscale eddy datasets.

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

This study utilizes gridded SSH and sea surface flow field data provided by the Surface Water and Ocean Topography (SWOT) mission to support the identification of submesoscale eddies based on satellite altimetry and to construct a corresponding eddy dataset. The SWOT satellite, jointly developed by the Centre National d'Études Spatiales (CNES), is dedicated to monitoring elevation changes of global water bodies, including ocean surfaces. Its primary objective is to enhance scientific understanding of the global water cycle, sea-level rise, extreme weather events, and climate change by delivering high-precision elevation data for both oceanic and terrestrial water surfaces (Fu et al., 2024). The satellite operates on a 21-day repeat orbit with a swath width of 120 km. Its Level-3 sea surface height product achieves a horizontal resolution of up to 2 km, endowing it with the capability to detect submesoscale oceanic eddies. SWOT data can be accessed at https://www.aviso.altimetry.fr/en/data.html. In addition, this study utilized gridded SSH fusion data with a spatial resolution of 0.125° × 0.125°to support global multi-scale eddy identification and to construct a global submesoscale eddy dataset. The gridded altimetry products are generated by the Copernicus Marine Environment Monitoring Service (CMEMS) through the fusion of data from multiple satellite altimetry missions, including SARAL/AltiKa, Cryosat-2, Jason-3, and Sentinel-3A & 3B (https://data.marine.copernicus.eu/products). The dataset enables daily global coverage, thereby providing a solid data foundation for multi-scale eddy identification, dataset construction, and eddy tracking. The Level-2 SST data employed in this study were acquired from the Moderate-Resolution Imaging Spectroradiometer (MODIS), featuring a spatial resolution of 1 km and available at https://oceancolor.gsfc.nasa.gov/data/10.5067/COMS/GOCI/L2/OC/. Within the temperature field, both the helical (or spiral) structures and associated thermal anomalies served as key indicators for identifying eddy formation. Table 1 provides a detailed overview of the data sources employed throughout this study.

Table 1. Production information used in this study.

Data	Satellite or instrument	Resolution	Level
Chlorophyll	Sentinel-3/GOCI-I	300/500 m	2
SST	MODIS	1000 m	2
SSH1	SWOT	2000 m	3
SSH2	Merged Data (multiple satellites)	0.125°	4

3 Methods

3.1 Submesoscale eddy identification based on deep learning algorithm and chlorophyll data

3.1.1 Chlorophyll image processing

For chlorophyll images provided by Sentinel-3A&B or GOCI, a single image can cover an extensive spatial area, with lateral dimensions exceeding a thousand kilometers. Consequently, such images are highly likely to encompass both coastal and open-ocean regions simultaneously, leading to substantial variations in observed chlorophyll concentrations within the same

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

image. This situation gives rise to two major issues. First, at the local scale, submesoscale eddies typically exhibit spatial extents on the order of kilometers to tens of kilometers, resulting in relatively small internal variations in chlorophyll concentration. In open ocean regions, where chlorophyll concentrations are inherently low, the chlorophyll concentration variations within submesoscale eddies are even smaller, thereby masking the chlorophyll spiral structure signals within eddy interiors and consequently hindering the identification of submesoscale eddies in the open oceans. Second, at the global scale—given the broad coverage of each chlorophyll image—a single scene may concurrently capture both nearshore and pelagic zones. As a result, adjusting the colorbar range of an image often enhances the visibility of spiral eddy structures in only one portion of the image, while signals in other regions remain indistinct. Moreover, the iterative process of manually tuning colorbar ranges is labor-intensive and time-consuming, rendering it impractical for large-scale data processing tasks. For example, the Sentinel-3 chlorophyll data used in this study comprise approximately 6,000 images per month from a single satellite, and over 10,000 images per month from the twin-satellite configuration, with an annual data storage requirement of about 20 TB. Therefore, to accurately detect chlorophyll spiral structures across different regions within countless large-scale remote sensing images, it is essential to simultaneously address two challenges: improving data processing efficiency and enhancing local features within chlorophyll images.

In this context, we proposed an image processing method specifically designed to enhance the spiral structures of chlorophyll in global oceans. The procedure is as follows: (a) To mitigate the influence of data noise, chlorophyll values were first constrained within the range of 0–100 mg/m^3 to remove outliers, thereby ensuring the reliability of subsequent analyses. (b) The chlorophyll data were then subjected to a base-10 logarithmic transformation, followed by normalization to scale the transformed values into the 0–1 range. Logarithmic transformation does not alter the correlation structure between data in different regions but compresses their range, reducing the impact of vast variations in chlorophyll concentration in different regions. Normalization, while preserving the correlation between chlorophyll data, further reduces the data range across the entire remote sensing image, preparing the data for subsequent processing. (c) Then, these normalized values were linearly stretched to a 0–255 scale. This step serves a dual purpose: it amplifies subtle local variations in oceanic chlorophyll concentrations, making the spiral structures of submesoscale eddies more discernible, and it standardizes the value range across all satellite images, eliminating the need for individual colorbar adjustments and significantly accelerating the image processing workflow. (d) Finally, the processed image data were enhanced using Contrast-Limited Adaptive Histogram Equalization (CLAHE) to further improve feature visibility. The CLAHE algorithm is a commonly used image enhancement technique in digital image processing, derived from the Adaptive Histogram Equalization (AHE) algorithm. AHE enhances image contrast by computing the image histogram and applying nonlinear transformations to expand the intensity values, thereby improving texture features. However, AHE has certain limitations, such as amplifying noise excessively in relatively uniform regions of the image. CLAHE addresses these drawbacks by dividing the image into smaller, localized regions, applying AHE to each region individually, and ultimately restricting the contrast enhancement within these areas (Vidhya et al., 2017; Reza et al., 2004). This approach makes local texture features more pronounced and mitigates the influence of the

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

overall chlorophyll concentration distribution across the entire image. Figure 1 illustrates the process of chlorophyll image processing.

Figure 2 illustrates the output results obtained using the traditional method of individually adjusting the colorbar value range and the approach proposed in this study. Figure 2b shows the result processed by the traditional method, where a small region of the satellite image is cropped, and the colorbar is adjusted to an appropriate range to emphasize the spiral structure of the eddy. Clearly, this method is highly inefficient when processing large-scale datasets. Additionally, the results are often unsatisfactory. As shown in Fig.2b – 2d, despite multiple adjustments to the color bar to achieve optimal outcomes, the image in Fig.2b remains less clear than those in Fig.2c–d, particularly in regions where chlorophyll concentration changes are minimal. Figure 2c displays the final result of the chlorophyll data processed using the method proposed in this section, while Fig.2d shows its pseudocolor image. It can be observed that the output images using the proposed method have a unified colorbar range (0–255), unaffected by regional variations in chlorophyll concentration. This greatly simplifies the data processing procedure and produces better results, with the spiral texture of the eddy becoming more pronounced and clearer.

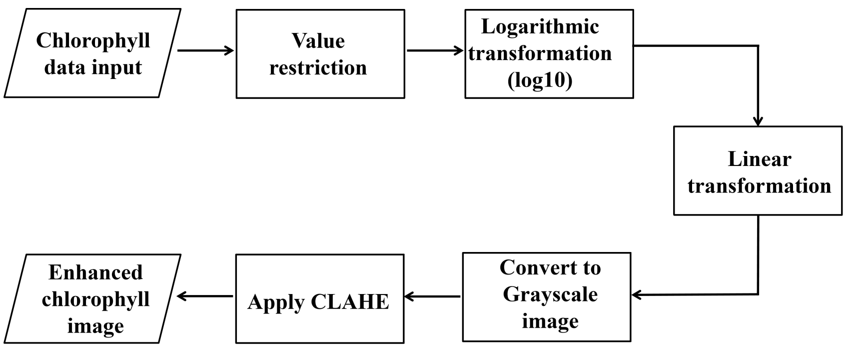

Figure 1. Chlorophyll data processing procedure

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

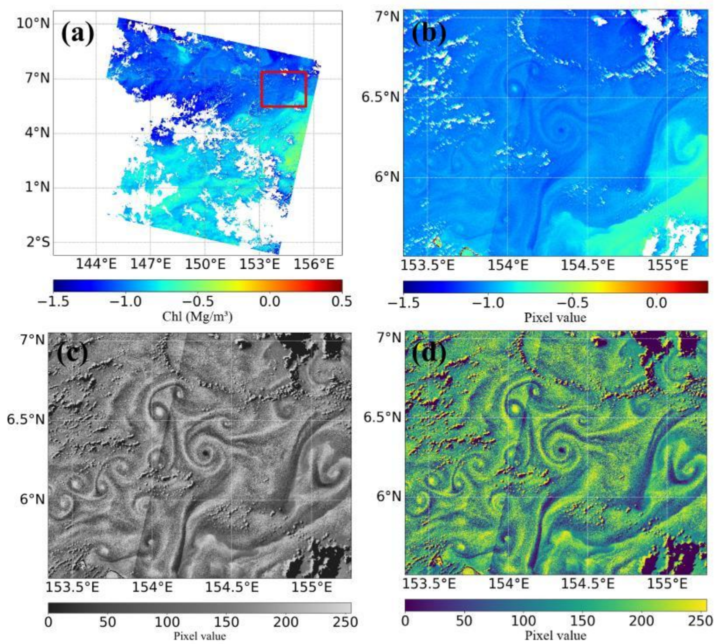

Figure 2. Eddy spiral structures shown in a satellite image provided by Sentinel 3B on March 17th, 2024. (a) Raw satellite image, with the red box indicating the geographical position of Figures (b)–(d). (b) Chlorophyll data processed by the traditional method (by refining the color bar scale for the local area in the image to display the eddy structures). (c) The final effect of chlorophyll data processed by the proposed method (all the images are accompanied by a unified color bar for consistent visualization). (d) The pseudo-color visualization of (c).

3.1.2 Deep learning method

To identify the spiral structures exhibited by submesoscale eddies in chlorophyll images, several key issues must be addressed. First, considering the size variations of submesoscale eddies, smaller eddies typically have radii on the order of kilometers, while larger eddies can extend to several tens of kilometers. Therefore, the task of eddy detection must account for the impact of these size differences, and the neural network model utilized should possess the capability for multi-scale object recognition. Second, some existing deep learning-based eddy detection methods only use rectangular bounding boxes to annotate the eddies shown in satellite images (Li et al., 2024; Xia et al., 2022; Wang et al., 2024). These methods overlook the irregular shape of eddies, which are typically elliptical in form. The use of square bounding boxes cannot accurately capture the contours of the eddies and the precise extent of their coverage. Therefore, it is necessary to delineate their boundary contours when detecting the submesoscale eddies. Third, eddies displayed in the chlorophyll field may exhibit a multi-layered spiral structure. If the receptive field of the neural network is too small for larger eddies, it may erroneously identify the inner layers as a complete eddy while overlooking the outer layers, leading to incomplete recognition of the eddy structure. On the other hand, smaller eddies face the opposite challenge. Therefore, enhancing the receptive field of the neural network to improve the accuracy of eddy detection is an urgent issue. Moreover, the marine environment is complex

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

and diverse, and the remote sensing images of chlorophyll fields reflect the intricacy of the background environment. The main task is to enable the model to better focus on the texture features of the eddies and accurately identify them amidst the complex marine background.

Generally, deeper convolutional layers in a neural network progressively emphasize global semantic features, including eddy scale and overall morphology. In contrast, the detailed local spiral structures that represent eddy features are more readily captured in shallower convolutional layers. Consequently, the selected neural network must be capable of effectively capturing shallow features while preserving deeper information. Against this background, DeepLabv3+ was chosen for the submesoscale eddy detection and segmentation mission (Chen et al., 2018). It can identify eddies in complex oceanic settings and precisely outline their contours. Figure 3 illustrates the detailed structure of DeepLabv3+. DeepLabv3+ is an improved version of the DeepLabv3 neural network, incorporating a decoder module to enhance image recognition accuracy and detail. It comprises two main modules: an encoder and a decoder. At the stage of encoding, the backbone network extracts image information using the Atrous Spatial Pyramid Pooling (ASPP) module. By parallelizing multiple atrous convolutions with varying dilation rates alongside global average pooling, the ASPP effectively captures multi-scale image features. During the decoding stage, upsampling is progressively applied to the encoded features. The low-level features from the backbone's earlier layers are then fused with the high-level features extracted by ASPP, enabling detailed reconstruction of image structures, particularly at boundaries, while preserving multi-scale target information. This method simultaneously retains the superficial and deep features of the eddy image, thereby improving the accuracy of submesoscale eddy recognition. In comparison to its counterparts, U-Net et al., DeepLabv3+ offers a more streamlined architecture while delivering superior accuracy in target identification (Dias et al., 2025). This architecture is widely used in medical image segmentation, autonomous driving, and remote sensing applications.

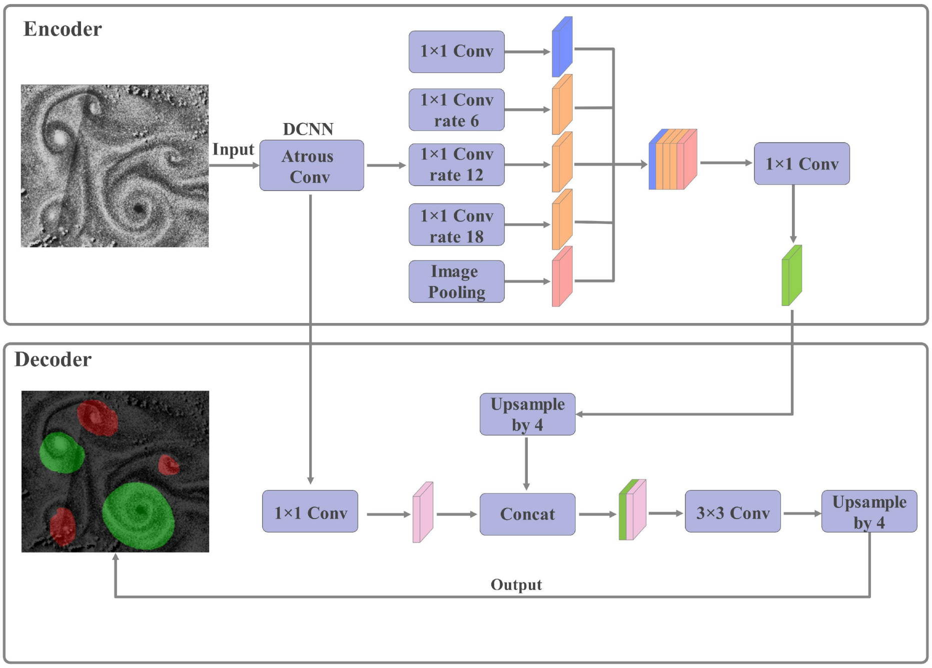

Figure 3. Overview of DeepLabv3+

3.1.3 Training set and implementation details

We employed three years of global chlorophyll data (from January 1, 2020, to December 31, 2022) from Sentinel-3A&B satellites to identify submesoscale eddies and create their dataset. The use of data with two different spatial resolutions enhances the robustness of the training outcomes. These data were first preprocessed using the proposed data processing method. Then, we randomly selected a subset of images from the downloaded data to form the training set for deep learning. After excluding images devoid of chlorophyll data, a total of 782 images were ultimately retained for constructing the training set, each containing several to dozens of eddies. Subsequently, CEs and AEs within these images were annotated using the LabelMe software. To balance the number of AEs and CEs samples, all remote sensing images were replicated and mirrored, yielding 6,966 CEs and 6,966 AEs. To further enhance the robustness and generalization capability of the eddy detection neural network and to mitigate limitations arising from insufficient diversity, image data augmentation was applied in each training epoch. Augmentation operations included scaling, rotation, and the addition of salt-and-pepper and Gaussian noise, enabling the model to adapt to submesoscale eddies with varying morphologies and textural characteristics. The dataset was randomly split into a 9:1 ratio, with 90% of the data used as the training set and the remaining 10% as the test set for evaluating eddy detection performance.

We used PyTorch 1.7.1 as the deep learning framework in this experiment to train the neural network model Deeplabv3+. The computational hardware consisted of an NVIDIA RTX A6000 GPU, with CUDA version 11.0 selected for acceleration.

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

Given the substantial memory requirements and the need to accelerate model convergence, we adopted the Adam optimizer, which enables automatic adjustment of model parameters. During training, the batch size was set to 4, with an initial learning rate of 0.0001 and a decay factor of 0.01 per epoch. A total of 200 training epochs were conducted until the loss function stabilized and ceased to decrease. The training results show that this neural network model performs well in the eddy detection and segmentation task. The Intersection over Union (IoU) reached 80.06%, the Precision reached 85.39%, and the Accuracy reached 93.49%. The performance is significantly better than some of the current cutting-edge research (Xu et al., 2023).

3.2 Submesoscale eddy identification method based on altimeter data

The method used for the eddy identification task based on SWOT and altimeter data refers to the topological method proposed by Chelton. (Chelton et al., 2011), and has been improved according to the requirements. By adding techniques such as vorticity verification, the accuracy and computational efficiency of eddy identification have been enhanced. The specific steps of the algorithm are as follows:

First, the altimetry data provided by SWOT were subjected to high-pass filtering to mitigate the effects of large-scale signals. Then parallel computing techniques were applied to process each data block simultaneously, further increasing computational speed. Assuming the original SLA image is divided into L regions, and the number of threads used for parallel computing is K, the computational complexity can be reduced from $O(n^2)$ to $O(n^2/(K \times L^2))$. The second step was the closed contour retrieval process. The search range was set from the minimum to the maximum SLA value within the selected region, with a height interval of 0.1 cm. This differs from the approach of Chelton et al., who used an interval of 1 cm across a range from –100 cm to 100 cm for mesoscale eddy identification (Chelton et al., 2011). Although a 1 cm interval is sufficient for detecting mesoscale eddies, a finer step size enables higher accuracy in detecting submesoscale eddies. Figure 4a illustrates closed contours identified in the SLA field. To highlight the diversity in the shapes and nesting structures of these closed contours, gridded data from the merged SLA product at a resolution of 0.125° were used for visualization. Within some closed contours, multiple extrema exist with nested structures, and the shapes and sizes exhibit considerable complexity and diversity. In traditional mesoscale eddy datasets, due to insufficient spatial resolution, these subtle variations in SLA data may not be resolved, leading to the loss of submesoscale signals. Therefore, extracting submesoscale eddy information from such complex, nested multilayer closed contours is a critically important task. The third step is eddy information extraction. After retrieving the closed contours, iterative processing should be performed on these contours to identify eddy boundaries, cores, polarity, and other attributes. The most critical step is the determination of eddy boundaries. In order to ensure the eddy identification results more accurate, the selection of closed contours must satisfy the following criteria: (a) If a closed contour contains only one extremum point and does not enclose any other contours, it should be retained. (b) If multiple nested closed contours share a common extremum, the outermost contour is retained. (c) If multiple nested closed contours contain multiple extrema, iterative processing is applied. A contour is retained if it contains only one extremum and is the outermost contour among all contours centered on that extremum. (d) The selected closed

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

contours must also pass a shape error test, with the shape error compared to the standard circle being no more than 55%. When a closed contour satisfies all the above criteria, the region enclosed by the contour is preliminarily identified as an eddy, and the extremum within the contour is taken as the eddy core. If the extremum point enclosed by the closed contour is a maximum, the eddy is classified as an AE; conversely, if it is a minimum, the eddy should be classified as a CE. The longitude and latitude coordinates of the eddy boundary and core indicate their corresponding geographic locations. Figure 4b shows some examples of detected eddies, where the blue curves represent CEs, and the red curves represent AEs. Fourth, in addition to distinct sea surface height anomalies, ocean eddies also exhibit important flow field characteristics. In general, AEs exhibit convergence, whereas CEs feature divergence. In the Northern Hemisphere, currents within AEs rotate clockwise with negative relative vorticity, while those inside CEs rotate counterclockwise with positive relative vorticity. The opposite pattern prevails in the Southern Hemisphere. Based on the above properties, we calculated the averaged relative vorticity within each detected eddy. Subsequently, only eddies with consistent polarity, flow direction, and vorticity were retained as valid submesoscale eddies, and those failing to satisfy all criteria were eliminated.

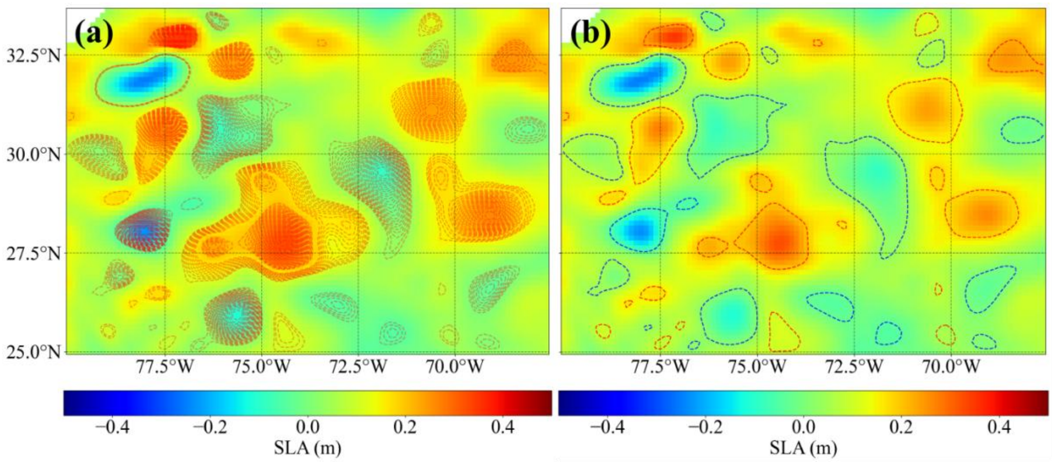

Figure. 4 (a) The closed contours in the SLA field, (b) the identified eddies in the SLA field.

4 Results and Discussion

4.1 Eddy identification

For chlorophyll data, we applied the trained neural network model DeepLabV3+ to detect submesoscale eddies within the chlorophyll images. Figure 5a presents several submesoscale eddies detected by the neural network model, where the red and green regions represent AEs and CEs, respectively. It can be observed that the model accurately identifies eddies of varying sizes and morphologies, demonstrating the effectiveness of this method for eddy detection. Furthermore, this eddy identification approach not only detects the presence of eddies but also clearly characterizes their polarity, coverage area, shape, and boundary information. As shown in Fig. 5a, each identified eddy appears as a contiguous region encompassing

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

the entire eddy structure. Before these data can be utilized, further processing is required to extract key information, such as the geographical coordinates of eddy boundaries and the eddy cores. The contour tracing algorithm proposed by Suzuki was employed to extract and record the pixel positions along the eddy boundaries (Suzuki et al., 1985). Since each pixel in the satellite image contains its own geographical coordinates information, the geographical coordinates of each eddy boundary point can be obtained by matching the eddy boundary pixel position with the corresponding pixel geographical information of satellite images. The eddy core was defined as the centroid of the closed contour of the eddy boundary. The area and radius of the eddy were then calculated based on the number of pixels enclosed by the eddy boundary and the spatial resolution of the satellite image. These eddies were classified into different polarities according to their rotation directions. In the Northern Hemisphere, eddies rotating counterclockwise were defined as CEs, while those rotating clockwise were defined as AEs. The opposite classification applied in the Southern Hemisphere. Figure 5b illustrates the final representation of the eddy data derived from chlorophyll fields, where the red closed curves denote AEs, and the blue closed curves denote CEs. In the end, chlorophyll-a data from January 1, 2020, to December 31, 2022, were downloaded for both Sentinel-3A and Sentinel-3B. Over 122,000 AEs and 1,282,000 CEs were identified. The submesoscale eddy information detected in each image was saved as an individual Pickle file, including their time, radius, polarity, as well as the coordinates of the eddy centers and boundaries. These data were organized into separate folders according to satellite platform, compressed into RAR archives, and uploaded to the Zenodo website. Users can access the data by simply decompressing the root folder, thereby obtaining all eddy information contained in the subfolders without requiring further processing. This dataset is available at the following website: https://zenodo.org/records/17483940 (Hou. 2025).

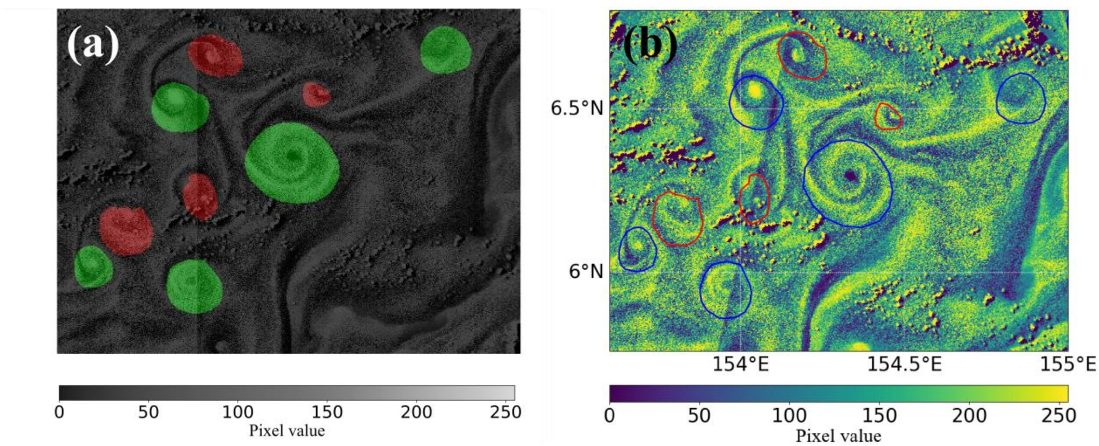

Figure 5. Oceanic eddies shown in chlorophyll fields. (a) Eddies detected in chlorophyll fields derived from Sentinel-3, with the red areas representing AEs, while the green areas represent CEs. (b) Eddy data with geographical coordinates. The red ellipses represent AEs, while blue ellipses represent CEs.

Figure 6 shows some submesoscale AEs and CEs identified from SLA data. The background fields consist of SLA and current velocity data derived from SWOT, with black arrows indicating current directions. Fig. 6a and 6b indicate a submesoscale CE and a submesoscale AE in the Northern Hemisphere, while Fig. 6c and 6d show those in the Southern

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

Hemisphere. As observed, eddies with opposite polarities rotate in reverse directions within the same hemisphere. For eddies of the same polarity, their rotational directions differ between the two hemispheres. The center of AEs features positive sea level anomalies, whereas CEs are characterized by negative anomalies at their centers. Currents within CEs present an outward divergent pattern, and those within AEs exhibit a convergent pattern. All these characteristics are consistent with well-documented properties of ocean eddies reported in existing studies (Chen et al., 2022). Notably, the relative vorticity inside each detected eddy was calculated throughout the identification procedure to verify the consistency between eddy detection results and vorticity signatures, which further improves the reliability of eddy identification. Finally, we constructed two datasets for eddies detected from two types of altimetry data. For SWOT satellite altimeter data, eddy information extracted from each satellite orbital dataset was stored in individual Pickle files. These files record the longitude and latitude of eddy centers and boundaries, as well as eddy radius, polarity, vorticity and other relevant parameters. The dataset covers a two-year period from July 26, 2023 to July 25, 2025, containing 6,045,154 AEs and 6,166,239 CEs. For eddies retrieved from the merged SLA data with a spatial resolution of 0.125°, daily global eddy records were archived in NetCDF files. The corresponding eddy dataset spans from January 1, 2012 to December 31, 2023, including 21,181,891 AEs and 22,109,575 CEs. Detailed descriptions and definitions of all parameters in the eddy datasets are available at the following website: https://zenodo.org/records/20714411 (Hou ,2026) and https://zenodo.org/records/20718880 (Hou et al., 2026).

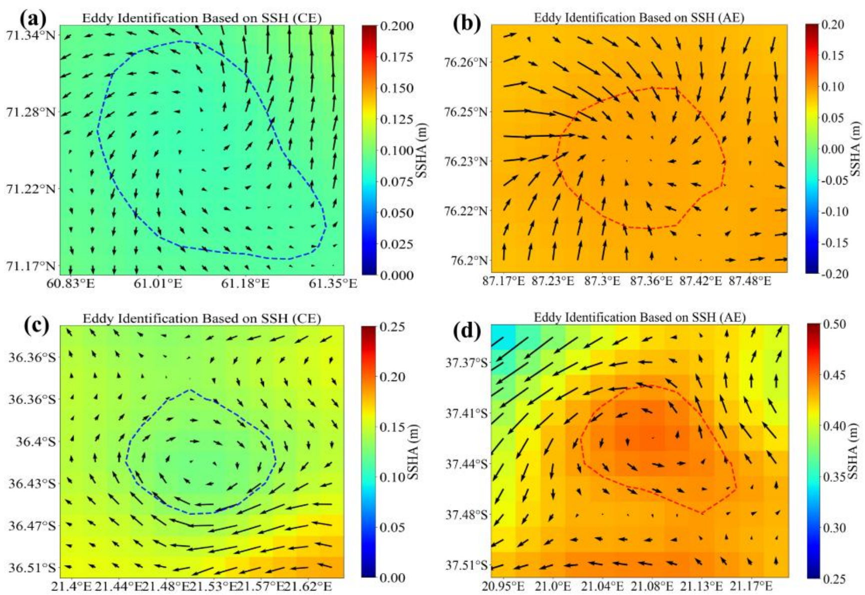

Figure.6 Some eddies identified in SLA data provided by SWOT. (a) and (b) are CEs and AEs in the Northern Hemisphere. (c) and (d) are CEs and AEs in the Southern Hemisphere. The background field is the sea surface height data, and the black arrows indicate the flow direction

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

SST anomalies and SSH anomalies induced by eddies are generally considered to be indicators of oceanic eddy presence. To evaluate the differences and similarities between the proposed submesoscale eddy identification method and the resulting datasets, we compared chlorophyll-a data, SST data, SLA data, and the mesoscale eddy dataset provided by AVISO. Specifically, we analyzed whether eddies identified from the three data sources exhibit temperature anomalies in the SST field, their differences, as well as the respective advantages and limitations of each dataset. Figure 7 shows the performance of some eddies identified from different datasets under the same spatiotemporal conditions, with detailed information provided in the caption. For ease of comparison, representative eddies in the figure have been numbered. For eddies identified from chlorophyll-a data, Fig. 7a−b show that eddies detected in the chlorophyll field correspond to clear spiral structures in the SST field at the same locations. For instance, eddy No. 4, despite having a radius of less than 10 km, still induces a noticeable temperature anomaly. This demonstrates both the effectiveness of chlorophyll-a data in submesoscale eddy identification tasks and the significant impact of submesoscale eddies on the environment. On the other hand, eddies No. 1 and No. 2 exhibit opposite polarities in both rotation direction and temperature anomaly. However, due to limitations in spatial resolution and eddy amplitude intensity, they were identified as a single eddy in the AVISO dataset and were not fully resolved even in the SLA data at 0.125° resolution. Thus, chlorophyll-a data, to some extent, compensate for the shortcomings of traditional datasets and play a crucial role in correcting and complementing conventional eddy datasets. As shown in Fig. 7c–d, for eddies identified from SLA data at a spatial resolution of 0.125°, the identification accuracy is higher than that of the AVISO mesoscale eddy dataset due to its improved spatial resolution. For example, in the AVISO mesoscale eddy dataset shown in Fig. 7c, the boundaries of three independent eddies were incorrectly merged and identified as a single large eddy. In contrast, the SLA data at 0.125° resolution in Fig. 7d clearly resolves them as separate eddies. Furthermore, the eddies identified from the SLA data at 0.125° resolution in Fig. 7b are also closer to those identified from the chlorophyll data. This indicates that the accuracy of eddy identification using altimetry data is closely related to the spatial resolution of the data. Figure 8 presents the eddy size distributions derived from the three datasets. Eddies identified from SWOT data exhibit the smallest radii, mostly within 8 km, and those shown in Fig. 6c−d are even smaller, with radii around 3 km. By comparing, we found that only those with relatively larger radii can also be detected in the chlorophyll data and the SLA data. The majority of smaller eddies, however, remain undetected by the latter two datasets. Therefore, from the perspective of spatial scale alone, SWOT data demonstrate the best performance for submesoscale eddy identification tasks.

It is worth noting that there are certain differences between chlorophyll data and altimetry data in eddy detection tasks. Eddies identified from chlorophyll data are not necessarily detectable by altimetry data, and conversely, eddies detected by altimetry data are not always identifiable in chlorophyll data. Some eddies, due to their low amplitude, may not be captured by altimetry data; however, their presence can still be inferred from the chlorophyll signatures induced by the stirring of water masses. On the other hand, eddies detected by altimetry may not be visible in chlorophyll images if the corresponding regions exhibit low chlorophyll concentrations or are obscured by cloud cover. Each type of data possesses its own advantages and characteristics in eddy identification tasks, and these datasets can complement each other in eddy detection.

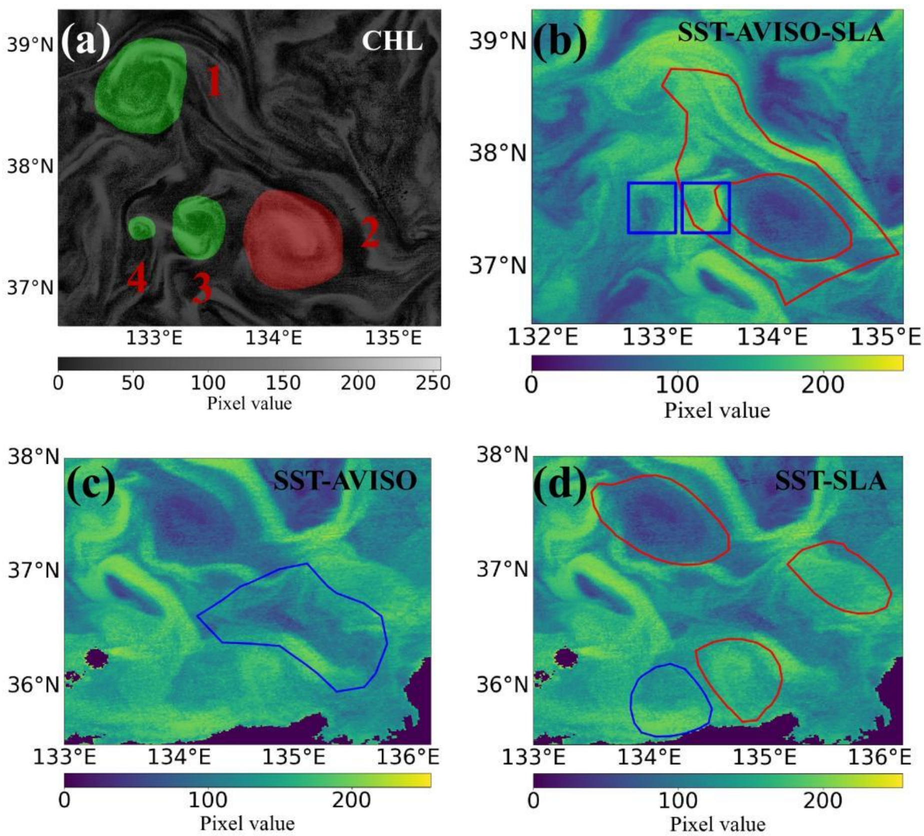

Figure 7. Oceanic eddies shown in several data sources. The background of (a) is chlorophyll fields, and the background of (b)-(d) are SST fields. (a) Eddies detected in chlorophyll fields derived from Sentinel-3, with the red areas representing AEs, while the green areas represent CEs. (b) Eddies detected in SLA fields. The bigger red ellipses represent one AE produced by AVISO, while the smaller one is an AE detected in the SLA field with a resolution of 0.125°. The blue boxes are two eddies shown in the SST field. (c) An CE produced by AVISO. (d) Eddies detected in SLA fields with a resolution of 0.125°, with red ellipses representing AEs and green ones representing CEs.

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

4.2 Statistical analysis of eddy characteristics

4.2.1 Scale Distributions

Scale characteristics are the most prominent features of submesoscale eddies, and the scale distribution of a submesoscale eddy dataset best reflects the effectiveness of the chosen data in the submesoscale eddy identification task. Specifically, the smaller the eddies that can be identified, the greater the advantage of the chosen data for this task.

Figure 8 illustrates the distribution of eddy scales identified by the three different datasets mentioned above. Fig. 8a and 8d display the eddy radii distribution detected using chlorophyll data provided by Sentinel-3A&B. The radii of eddies in both the Northern and Southern Hemispheres exhibit striking similarities, with the majority concentrated within a radius of 25 km. Both AEs and CEs tend to aggregate toward smaller scales, peaking within the 5−10 km range. The average radius of AEs and CEs in the Northern Hemisphere are 12.46 km and 10.54 km, respectively, while in the Southern Hemisphere, they are 12.17 km and 11.18 km, respectively. Compared to the global daily eddy dataset with a resolution of 0.25°, the chlorophyll data allows for the identification of smaller-scale eddies while maintaining complete global coverage on a daily basis. This results in a distinctive advantage in the identification and tracking of submesoscale eddies. Globally, the average radius of AEs is significantly greater than that of CEs, a phenomenon that mirrors the differences in scale radii of mesoscale eddies. In terms of quantity, CEs are far more abundant than AEs, marking a notable distinction in the eddy identification results between chlorophyll data and traditional altimetry data. This phenomenon will be explained when analyzing the eddies' geographical distribution later.

Figure 8b and 8e display the eddy scale distribution detected using sea surface altimetry data from SWOT, arranged in ascending order by radius with 2 km increments. AEs and CEs exhibit a high degree of similarity in their scale distributions, with a general trend indicating that smaller scales correspond to a higher number of eddies. The majority of eddy radii are concentrated within 10 km, peaking within the 2–4 km range, with average radii of 4.57 km for AEs and 4.94 km for CEs. In terms of both quantity and scale, AEs and CEs show little difference, demonstrating a balanced distribution that reflects the dynamical equilibrium of the ocean. The average radius being less than 5 km indicates that the satellite altimetry data provided by SWOT has a greater advantage in identifying submesoscale eddies compared to chlorophyll data. Figure 8c and 8f illustrate the eddy scale distribution identified using merged altimetry data with a resolution of 0.125°, arranged in ascending order by radius in 10 km increments. It is evident that the radii of AEs and CEs predominantly range between 10 and 70 km, peaking in the 20–30 km interval, with average radii of 40.7 km for AEs and 39.1 km for CEs. Compared to the SWOT and chlorophyll data, the merged altimeter data are less effective in identifying submesoscale eddies; however, they demonstrate superior performance in the identification of cross-scale eddies with radii between 25–50 km.

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

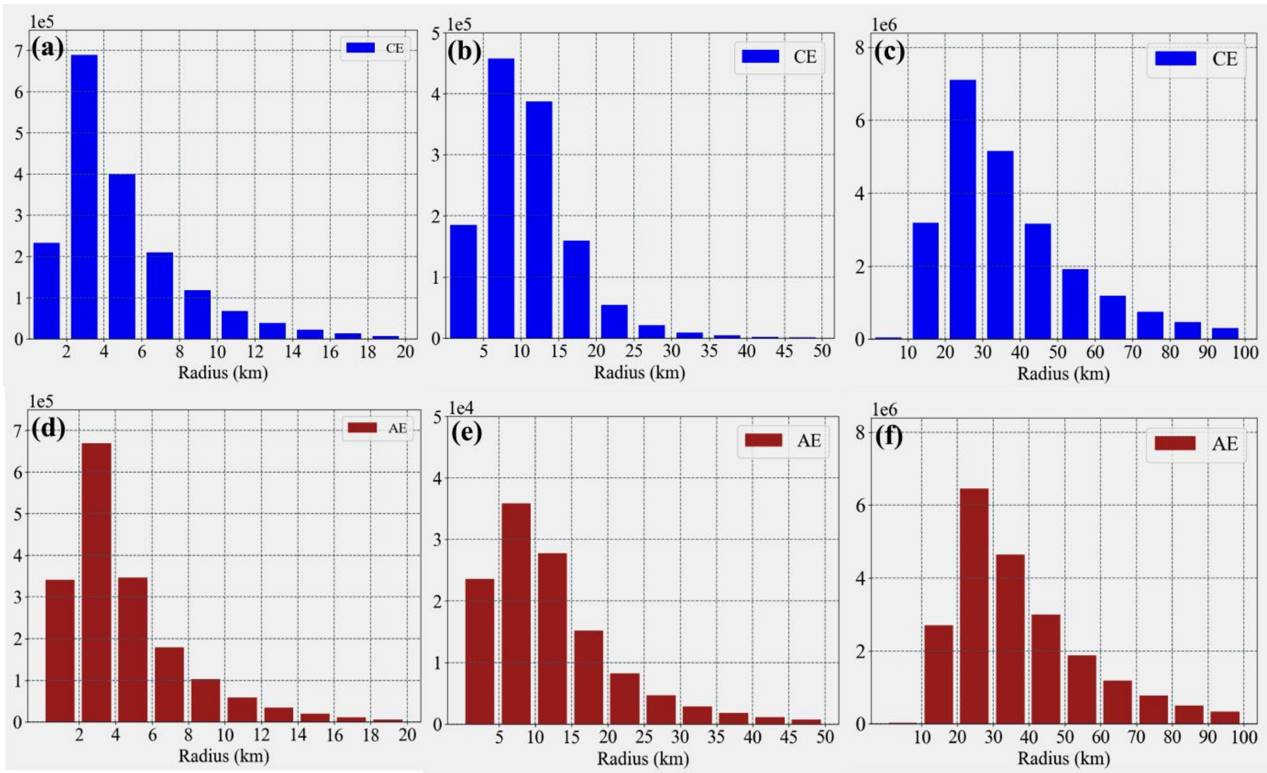

Figure. 8 The scale distribution characteristics of CEs and AEs identified using different data sources. (a) and (d) Submesoscale CEs and AEs identified by the SWOT satellite sea surface height data. (b) and (e) Submesoscale eddies identified using chlorophyll data and deep learning algorithms. (c) and (f) CEs and AEs identified from merged sea surface height data with a resolution of 0.125°

4.2.2 Geographical Distribution

Geographical features are one of the critical characteristics of submesoscale eddies. Influenced by factors such as ocean currents, topography, and formation mechanisms, the distribution of eddies varies across different regions. Furthermore, the geographical distributions of eddies identified by different datasets are uniquely affected by satellite resolution and meteorological environment conditions. This study, based on the three established eddy datasets, represents the first statistical analysis of the geographical distribution characteristics of submesoscale eddies at high temporal and spatial resolutions on a global scale. Figure 9 illustrates the geographical distribution characteristics of submesoscale eddies as detected using various datasets and methods, portraying the quantity of submesoscale eddies within every 2° × 2° grid cells globally.

From Fig. 9a–b, it is evident that the quantity of submesoscale eddies detected using chlorophyll data varies significantly across different regions of the globe, exhibiting clear regional characteristics. The distribution of these eddies in the global oceans is closely associated with their formation mechanisms, with a tendency to occur in areas of active ocean currents or regions characterized by complex sea conditions and topography (Yamada et al., 2004). Examples include the coastal

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

regions of Japan, the eastern coast of the Philippines, and the west coast of the United States, as well as active current regions such as the western waters of Australia, the Kuroshio region, subtropical circulation areas, and the Gulf Stream region. This correlation indicates that horizontal shear induced by abrupt changes in active ocean currents promotes the generation of submesoscale eddies (McWilliams et al., 2016; Roemmich et al., 2009). It is worth noting that the identified submesoscale eddies are relatively scarce in high-latitude regions, particularly in polar areas, where detecting such eddies poses significant challenges primarily due to optical conditions, climatic factors, and satellite viewing angles. Firstly, the occurrence of polar night during winter in both the Arctic and Antarctic leads to poor illumination and extensive cloud cover, making it difficult for passive optical remote sensing satellites to capture ocean color information. Moreover, the extensive ice cover in these polar regions, coupled with harsh low-temperature environments, creates unfavorable conditions for the survival of phytoplankton at the sea surface, resulting in a limited amount of submesoscale eddy data derived from chlorophyll measurements. Additionally, the low solar elevation angle in polar regions means that even during the summer months when daylight is present, the sun remains at a very low angle. This situation contributes to insufficient light intensity, resulting in limited solar radiation energy reaching the ocean surface, which leads to extremely weak signals received by sensors and poor signal-to-noise ratios. Furthermore, the prolonged path length of solar radiation through the atmosphere amplifies atmospheric scattering effects, causing the signals to be nearly obscured by strong atmospheric noise. Consequently, extracting weak marine reflection signals becomes exceedingly challenging. As a result, the geographic distribution of eddies identified from chlorophyll data indicates a notably low number in polar regions, with some areas even lacking any detection of such eddies altogether. Looking forward, investigating ocean eddies through sea surface ice detection or enhancing the signal-to-noise ratios of satellites operating under extremely low light conditions could represent promising avenues for future research.

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

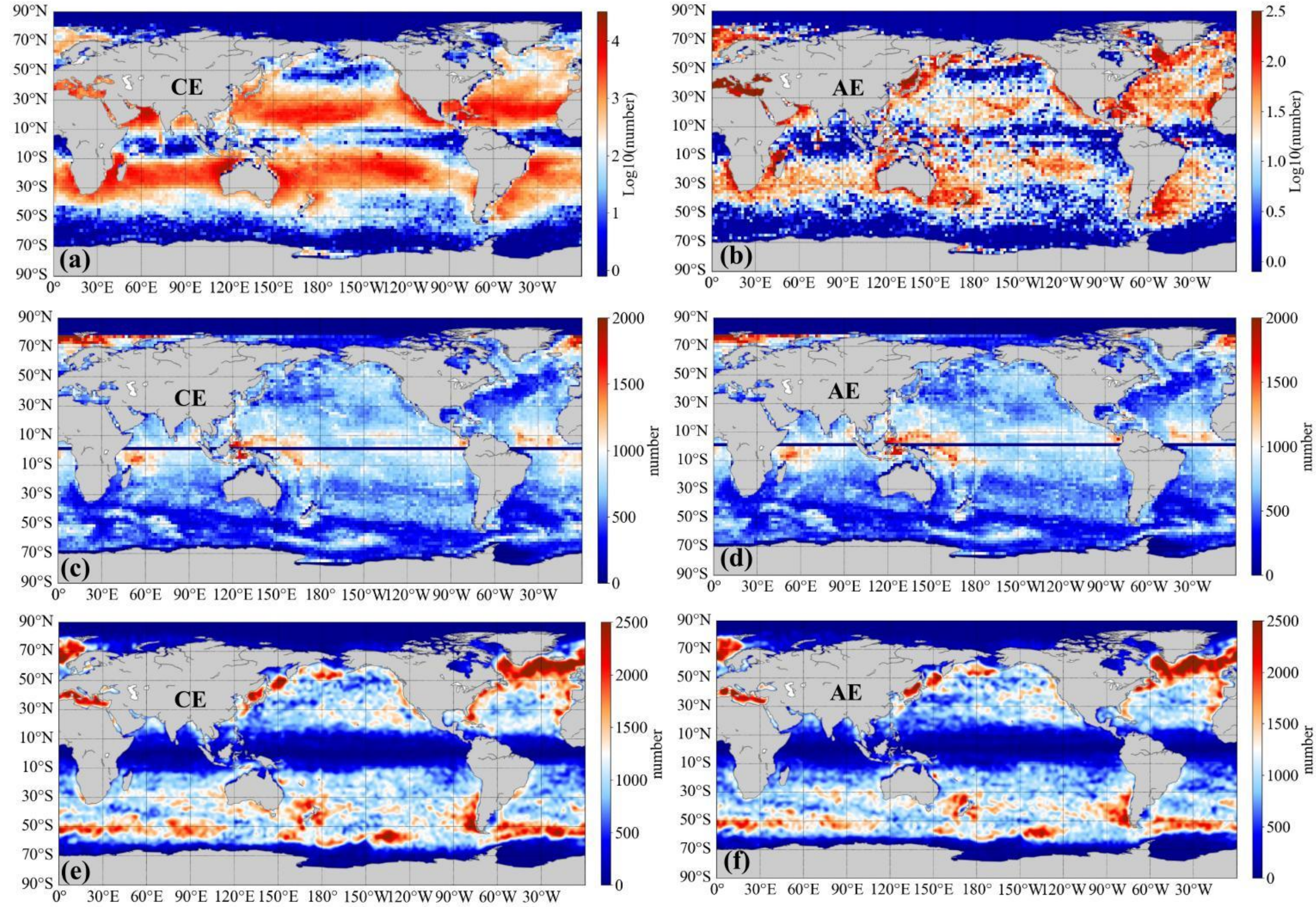

Figure. 9 Geographic distribution of global eddies from different datasets. (a, b) Geographic distribution of submesoscale CEs and AEs with radii less than 25 km derived from chlorophyll data. (c, d) Geographic distribution of submesoscale CEs and AEs with radii less than 25 km based on sea surface height data from SWOT. (e, f) The year averaged geographic distribution of CEs and AEs with radii less than 100 km obtained from merged altimetry data

Figure 9c and 9d show the geographical distribution of CEs and AEs with radii within 25 km, detected from SSH data provided by SWOT within each 2° × 2° grid. The number of submesoscale eddies varies considerably across different regions, exhibiting an overall trend similar to that of eddies identified from chlorophyll data. In terms of latitude, submesoscale eddies are mostly concentrated in mid- and low- latitude latitude regions, a pattern that differs from existing statistics of mesoscale eddy datasets, where mesoscale eddies are relatively scarce in low latitude belts. Regionally, submesoscale eddies are predominantly found in areas characterized by complex topography and active ocean currents, such as coastal regions and subtropical gyres. The horizontal shear ocean currents are more conducive to the generation of submesoscale eddies. Consequently, in mid- and low- latitude regions with active currents, although mesoscale eddies are relatively few in number, a substantial number of submesoscale eddies are present.

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

Due to the inclusion of numerous mesoscale eddies in the results identified from the merged altimeter data at a resolution of 0.125°, their distribution characteristics differ somewhat from those of eddies identified using the previous two data types. Specifically, this distribution more closely resembles that of mesoscale eddies provided by AVISO, with higher counts in mid-to-high latitudes and lower counts in low-latitude regions. In contrast, the selected eddies from the SWOT altimeter and Sentinel-3A&B chlorophyll are submesoscale eddies, resulting in markedly different geographical distribution characteristics compared to the former. Submesoscale eddies identified from these datasets are predominantly concentrated in mid-to-low latitudes, with fewer occurrences in high-latitude regions. This study reveals a clear discrepancy in geographical distribution between submesoscale eddies and conventional mesoscale eddies. This discrepancy arises because submesoscale eddies are often capable of breaking geostrophic constraints, which accounts for the abundant presence of submesoscale eddies even in low-latitude regions.

4.2.3 Seasonal Variations

Figure 10 illustrates the monthly average of detected submesoscale eddies in the Northern and Southern Hemispheres. Fig. 10a – b depict the seasonal variations of eddies identified from chlorophyll data, while Fig. 10c – d show the seasonal distribution of eddies identified from SSH data provided by SWOT. Figure 10e–f represent the eddy seasonal variations from fused altimeter data. It is evident that the eddies identified from chlorophyll data exhibit significant seasonal variations, whereas the seasonal differences in eddies identified from altimeter data are less pronounced.

For chlorophyll data, the total number of recognized submesoscale eddies shows a clear seasonal pattern, with a greater abundance during winter and spring and a reduced presence in summer and fall. On the one hand, the enhanced wind stress during winter increases the mixed layer depth, allowing for greater potential energy storage in subsurface submesoscale fronts. The instability of these fronts contributes to the generation of more submesoscale eddies in winter, leading to a higher eddy count during this season. This phenomenon has been noted in several studies, although this impact is relatively limited on a global scale (Schubert et al., 2020; Gula et al., 2022; Pan et al., 2023). The primary factor for the significant differences in the number of eddies identified from chlorophyll data can be attributed to changes in phytoplankton abundance caused by climate change. During winter and spring, phytoplankton growth elevates chlorophyll concentrations, making it easier to observe eddies, which results in a higher count of identified eddies. With the combination of oceanic physical environment and ecological factors, there are considerable seasonal differences in submesoscale eddies. It can also be seen that the number of CEs far exceeds that of AEs, a discrepancy that can be attributed to three main factors. First, centrifugal instability, frontal instability, and horizontal shear can lead to the formation of cyclonic shear and spirals, which makes submesoscale eddies more likely to manifest as cyclonic structures (Buckingham et al., 2017; DiGiacomo et al., 2001). A similar case has been observed in the Southern California Bight, where DiGiacomo reported that approximately 90% of the observed eddies in this area were CEs (Shcherbina et al., 2013). Second, compared to submesoscale AEs, submesoscale CEs exhibit greater intensity and larger Rossby numbers, resulting in more pronounced effects on ocean tracers. This characteristic makes them easier to identify in satellite imagery (Zhang et al., 2016). Third, CEs can transport nutrient-rich water from the ocean depths to the surface layers through vertical motion, providing essential nutrients for the photosynthesis of phytoplankton in the upper mixed layer, thus promoting phytoplankton growth. Consequently, the chlorophyll signals within CEs become more pronounced and are more easily detected by ocean optical satellites. Figure 11 presents the global distribution of the ratio of CEs to AEs, calculated as the number of CEs divided by the number of AEs within each 2° × 2° grid cell. This value reveals the magnitude of the numerical difference between CEs and AEs across different regions. It can be observed that the largest difference between CEs and AEs occurs in the open ocean areas of the mid- low latitude regions, while it is relatively smaller in other regions. Given that the vertical transport capacity of submesoscale eddies is substantially stronger than that of mesoscale eddies, in oligotrophic oceanic regions such as subtropical gyres, CEs can transport nutrients from the ocean interior to the surface, thereby promoting phytoplankton growth (Levy et al., 2018; Erickson et al., 2020; Pan et al., 2023). The enhanced phytoplankton biomass makes CEs more easily detectable. Consequently, the numerical disparity between AEs and CEs is more pronounced in open oceans.

For eddies detected from sea surface height data, substantial differences exist compared with those identified from chlorophyll data, both in total number and in seasonal distribution (Fig. 10c–d). In terms of quantity, CEs are slightly more numerous than AEs, but the difference is not pronounced. Seasonally, the variation is also relatively modest, with slightly higher counts in autumn and winter, and the overall trend remains fairly stable. Enhanced winter wind stress deepens the mixed layer, thereby allowing more available potential energy to be stored beneath the sea surface in submesoscale fronts. The instability of these fronts promotes the generation of more submesoscale eddies in winter, thus increasing the number of submesoscale CEs (Schubert et al., 2020; Gula et al., 2022; Pan et al., 2023). On the other hand, the identification of eddies using satellite altimetry data is carried out by calculating sea surface height anomalies. As a result, the environmental changes driven by seasonal variability do not lead to as pronounced a difference in the number of AEs and CEs as observed in chlorophyll data.

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

Open Access Earth System Science Data Discussions

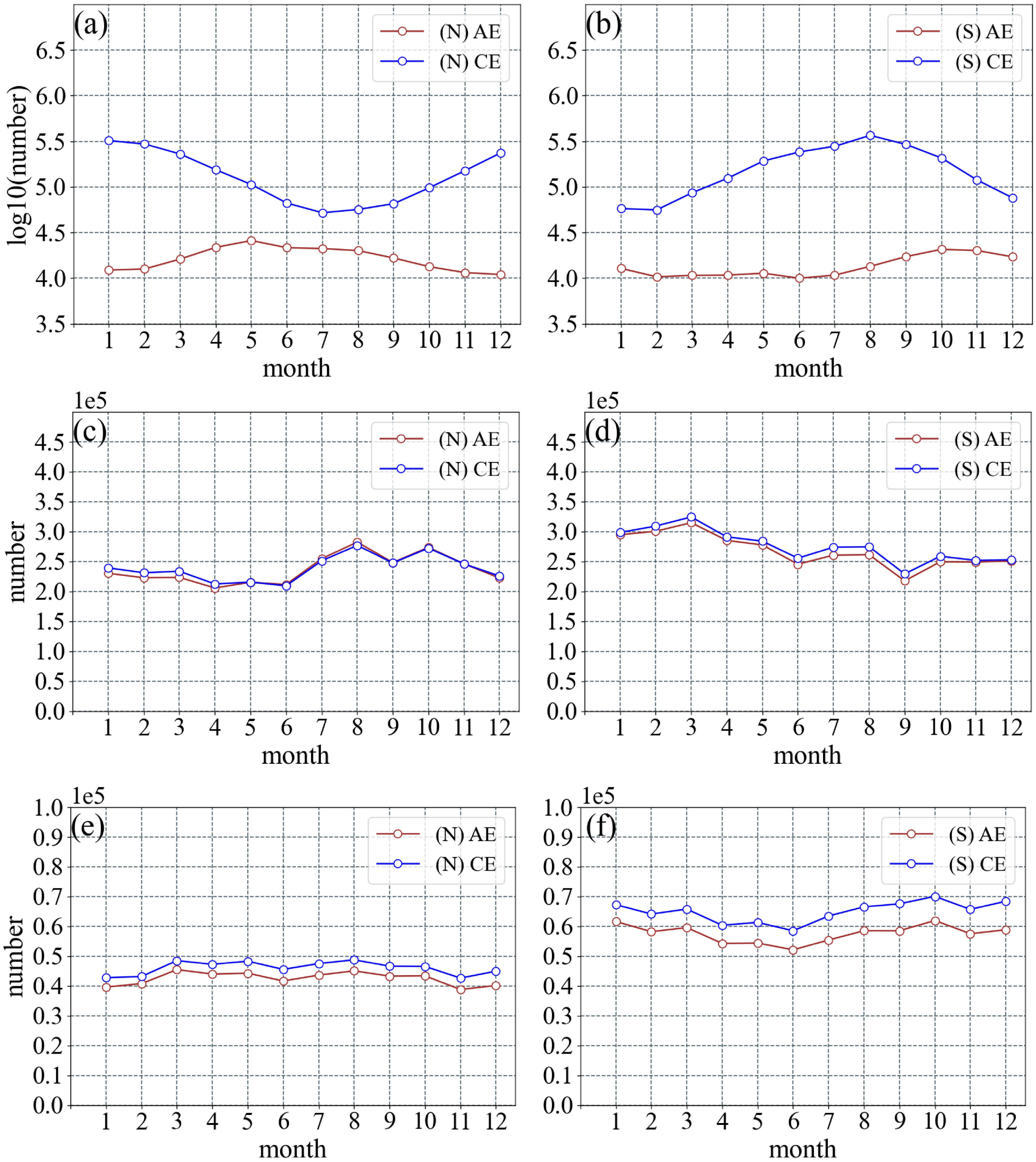

Figure. 10 Seasonal variation of submesoscale eddies. (a) and (b) represent the seasonal variation of submesoscale eddies in the Northern Hemisphere and the Southern Hemisphere derived from chlorophyll fields, respectively; (c) and (d) represent the year averaged seasonal distribution of the eddies identified through SWOT analysis in the Northern Hemisphere and the Southern Hemisphere; (e) and (f) represent the year averaged seasonal distribution of the eddies identified through merged SLA data in the Northern Hemisphere and the Southern Hemisphere. The blue, red, and black lines respectively represent the number of CEs and AEs

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

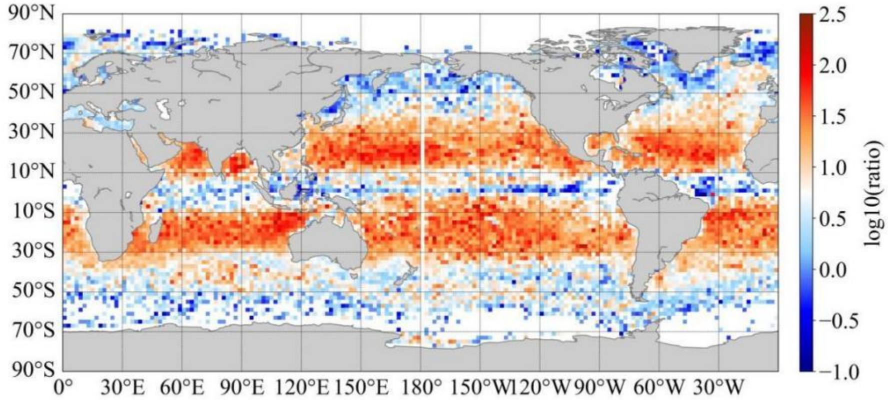

Figure. 11 The geographical distribution of the ratio of the quantities of CEs and AEs identified from chlorophyll field

4.3 Strengths and limitations of multi-source datasets

Nowadays, with diversified data sources and methodologies, a single data source or algorithm can hardly achieve satisfactory performance in ocean eddy identification. Each type of data possesses distinct strengths and limitations in eddy detection tasks. Based on the three established eddy datasets, we analyzed the advantages of various data and methods applied to eddy recognition, with a particular focus on submesoscale eddies that face a scarcity of available datasets. This study enhances the understanding of eddy identification mechanisms and facilitates the rational selection of data and algorithms for eddy research.

From Figure 8, it can be observed that the radii of eddies identified from sea surface height data provided by SWOT are predominantly below 10 km, with a concentration around 2–6 km. In contrast, the radii of eddies identified from chlorophyll data are mostly below 30 km, concentrating around 5–20 km. The eddy radii from multitask fusion sea surface height data with a resolution of 0.125° range between 10 and 70 km, with a focus on 20–50 km. Clearly, in terms of submesoscale eddy identification tasks, SWOT satellite data and chlorophyll data exhibit greater advantages, with SWOT satellite altimetry data capable of identifying the smallest eddy scales, followed by chlorophyll data. The fusion of sea surface height data with a resolution of 0.125° is more suitable for identifying cross-scale eddies with radii between 20 and 50 km, effectively bridging the gap between the mesoscale eddy data provided by AVISO and submesoscale eddies. In terms of spatial-temporal coverage, submesoscale eddies identified from chlorophyll data are significantly influenced by cloud cover and seasonal variations, resulting in data gaps during adverse weather conditions or in high-latitude regions. Although SWOT altimetry data are not affected by weather or seasonal changes, its limitations in temporal resolution pose challenges for studies on eddy tracking and the characterization of changes throughout their life cycle. While the chlorophyll data provides global coverage on a daily basis, is crucial for research on eddy tracking, the evolution of eddy spiral structures and shapes, as well as the ecological processes from formation to dissipation.

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

In summary, for submesoscale eddy identification task, the altimetry data provided by the SWOT satellite and the chlorophyll data from Sentinel-3 yield the best results, with SWOT altimetry data being superior. However, SWOT altimetry data faces challenges in submesoscale eddy tracking due to its temporal resolution limitations. In contrast, ocean color data represented by Sentinel-3 and GOCI, characterized by high temporal resolution and extensive spatial coverage, not only effectively identify submesoscale eddies but also facilitate their tracking and the study of their ecological impacts throughout their lifecycle, despite being significantly affected by cloud cover and weather conditions. Furthermore, the fusion height data with a resolution of 0.125° is better suited for identifying cross-scale eddies that fall between mesoscale and submesoscale dimensions.

5 Conclusion

This study integrates multi-source heterogeneous oceanic data, numerical computation, and artificial intelligence techniques to investigate submesoscale eddy detection and their characteristic distributions. Based on ocean color data and artificial intelligence technology, we designed a submesoscale eddy identification method and established a global submesoscale eddy dataset . Then, we developed an eddy identification method based on altimetry data, and constructed two separate eddy datasets: one from SWOT satellite altimetry data and the other from gridded altimetry data at a spatial resolution of 0.125°. We finally conducted statistical analyses on the eddy characteristics, including their scale, geographical distribution, and seasonal variation across the three datasets, and discussed the strengths and limitations of different data sources and methods for submesoscale eddy detection tasks. The detailed procedures and major findings are presented as follows:

We first developed a submesoscale eddy detection method using deep learning and digital image processing techniques combined with sea surface chlorophyll data, and constructed a global submesoscale eddy dataset with high spatiotemporal resolution. The chlorophyll image enhancement approach enables clear visualization of faint spiral chlorophyll structures. It eliminates manual color-scale tuning for individual chlorophyll images and avoids interferences induced by spatial variations in chlorophyll concentration across diverse regions, substantially improving eddy visualization and accelerating data processing efficiency. Experimental results demonstrate that the presented detection algorithm can reliably identify submesoscale eddies of diverse shapes and sizes within chlorophyll fields and delineate their outer contours simultaneously. Geographical information such as the longitude and latitude of each eddy was determined by mapping pixel coordinates of extracted eddy outlines onto geographic projections. This method overcomes key limitations inherent to conventional numerical approaches and low-resolution altimeter data, including rigid threshold selection, missing detections, and erroneous merging of discrete individual eddies. It provides critical support for validating and supplementing existing eddy datasets. Using this algorithm, we generate a fthree-year global submesoscale eddy dataset spanning from January 1, 2020, to December 31, 2022.

Based on satellite altimeter data, we developed an automatic identification algorithm for global submesoscale eddies. This algorithm improves computational efficiency by employing data partitioning and multi-threading techniques. It addresses the

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

issue of multiple extrema through iterative computation of complex closed contours and enhances the accuracy of eddy identification by combining closed contour analysis with vorticity calculation. The method partitions sea surface height anomaly data into blocks to reduce the number of iterations required for closed contour detection, thereby improving efficiency. Additionally, parallel computing techniques are applied to process each data block simultaneously, further increasing computational speed. Assuming the original SLA image is divided into L regions, and the number of threads used for parallel computing is K, the computational complexity is reduced from $O(n^2)$ to $O(n^2/(K \times L^2))$. The algorithm first determines eddy polarity and boundaries based on height anomalies and closed contours, and then validates the results by calculating the relative vorticity inside the eddy, leading to more accurate and reliable identification. Results show that the proposed eddy identification algorithm can accurately and effectively detect submesoscale eddies within the complex background of ocean height fields. Finally, using this method with SLA data from SWOT, we compiled submesoscale eddy statistics over a two-year period from July 26, 2023 to July 25, 2025, identifying 6,045,154 AEs and 6,166,239 CEs. Furthermore, based on merged altimeter data at a resolution of 0.125°, we derived eddy statistics from January 1, 2012, to December 31, 2023, identifying 21,181,891 AEs and 22,109,575 CEs.

Based on the three types of established eddy datasets, we conducted a characteristic analysis and comparison of submesoscale eddies identified from chlorophyll data and altimeter data. The results indicate that submesoscale eddies identified from sea surface chlorophyll data provided by Sentinel-3 using deep learning algorithm predominantly exhibit radii below 30 km, with most concentrated in the range of 5–20 km, demonstrating a clear advantage in the field of submesoscale eddy identification. Furthermore, in terms of geographical distribution, these eddies are mainly concentrated in mid-to-low latitudes, coastal regions, and areas with active ocean currents. The number of eddies in the Northern and Southern Hemispheres is significantly influenced by seasonal variations, primarily due to the effects of polar nights, cloud cover, and ice coverage in the polar regions. As the Sentinel-3A/B satellites enable global daily coverage, they exhibit a distinct advantage in submesoscale eddy tracking tasks. Submesoscale eddies identified from SLA data provided by SWOT predominantly exhibit radii below 10 km, with most concentrated in the range of 2–6 km. Among all datasets, these eddies have the smallest detectable scales. Therefore, in terms of spatial scale, SWOT satellite altimeter data are the most advantageous for submesoscale eddy identification. Geographically, these eddies are also concentrated in mid-to-low latitudes, coastal regions, and areas with active ocean currents. Compared with chlorophyll data, which are greatly affected by illumination conditions, the geographical distribution of submesoscale eddies identified from SWOT altimeter data is less influenced by seasonal variations. However, due to the satellite's 21-day repeat cycle, tracking submesoscale eddies with short lifespans and small scales using SWOT data presents certain challenges. Eddies identified from the merged altimeter data at a resolution of 0.125° exhibit radii ranging from 10 to 100 km, with most concentrated between 25 and 50 km, showing moderate performance in submesoscale eddy identification. Nevertheless, this dataset demonstrates unique advantages in studying cross-scale eddies that bridge the submesoscale and mesoscale ranges.

This study successfully accomplished the identification, statistical analysis, and evaluation of submesoscale eddies using multi-source data and a variety of methodologies. On the one hand, it will significantly augment, correct, and enhance the existing global ocean eddy datasets, thereby providing crucial support for the investigation of submesoscale eddy characteristics and their impacts on marine environments and fisheries. On the other hand, by analyzing the strengths and limitations of each eddy dataset in the context of submesoscale eddy identification tasks, as well as their respective applicability, this research fosters a deeper understanding of submesoscale eddies and offers valuable insights for data selection in eddy studies across different tasks.

Data availability

All the datasets established in the article can be freely accessed on the Zenodo website. The data constructed by Chlorophyll data, SWOT SLA data, and merged SLA data are aviliable at https://zenodo.org/records/17483940 (Hou, 2025), https://zenodo.org/records/20714411 (Hou, 2026), and https://zenodo.org/records/20718880 (Hou et al., 2026).

Author contributions

Meng Hou completed the investigation, method, formal analysis, software, visualization, and original draft writing of this manuscript. The supervision, project administration, and funding acquisition were supported by Jie Yang and Ge chen. Nan Luo reviewed and edited the manuscript.

Competing interests

The contact author has declared that neither of the authors has any competing interests.

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Financial support

This research was funded by the National Natural Science Foundation of China (No. 42530404) and the Laoshan Laboratory Science and Technology Innovation Projects (No. LSKJ202201302).

References

Buckingham, C. E., Z. Khaleel, A. Lazar, A. P. Martin, J. T. Allen, A. C. Naveira Garabato, A. F. Thompson, and C. Vic,: Testing Munk's hypothesis for submesoscale eddy generation using observations in the North Atlantic, J. Geophys. Res. Oceans, 122, 6725–6745, doi:10.1002/2017JC012910, 2017.

Chaigneau, A., Gizolme, A., Grados C.: Mesoscale eddies off Peru in altimeter records: Identification algorithms and eddy spatio-temporal patterns[J]. Progress in Oceanography, 79(2-4): 106-119, https://doi.org/10.1016/j.pocean.2008.10.013 Get rights and content, 2008.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 801-818). https://link.springer.com/conference/eccv, 2018.

Chen, G., X. Chen, and C. Cao.: Divergence and Dispersion of Global Eddy Propagation from Satellite Altimetry. J. Phys. Oceanogr., 52, 705–722, https://doi.org/10.1175/JPO-D-21-0122.1, 2022.

Chen, J., Yang, J., Tao, R., & Yu, Z.: Mesoscale eddy detection and edge structure extraction method in sar image. IOP Conference Series Earth and Environmental Science, 237 (3), 032010. https://doi.org/10.1088/1755-1315/237/3/032010, 2019.

Chelton, D. B., M. G. Schlax, R. M. Samelson, and R. A. de Szoeke,: Global observations of large oceanic eddies, Geophys. Res. Lett., 34, L15606, doi:10.1029/2007GL030812, 2007.

Chelton, D. B., Schlax, M. G., Samelson, R. M.: Global observations of nonlinear mesoscale eddies, Progress in Oceanography, 91(2): 167-216, https://doi.org/10.1016/j.pocean.2011.01.002, 2011.

Chelton, D. B, Gaube, P., Schlax, M. G.: The influence of nonlinear mesoscale eddies on near-surface oceanic chlorophyll. Science, 334, 328–332, DOI:10.1126/science.1208897, 2011b.

Clerc, S., Donlon, C., Borde, F., Lamquin, N., Hunt, S. E., Smith, D., ... & Guérou, A.: Benefits and lessons learned fr``om the Sentinel-3 tandem phase. Remote Sensing, 12(17), 2668. https://doi.org/10.3390/rs12172668, 2020.

Dias, Jean & Pereira, Williane & Negreiros, Waldemiro & Vale Guimarães, Pedro & Soares Corrêa, Alan & Tamasauskas, Leonardo & Gomes, Marivan & Souza, Danilo & Yoshida, William & Valente, Danilo & Rocha, Daniela & Augusto, Fernando & Brito-Costa, Gabriel & Seruffo, Marcos.: Cloud Segmentation in Multispectral Images from the Sentinel-2 Satellite: A Comparative Study of Deep Learning Approaches. 59-68. 10.5753/wcama.2025.7754, 2025.

DiGiacomo, P. M., and B. Holt,: Satellite observations of small coastal ocean eddies in the Southern California Bight, J. Geophys. Res., 106(C10), 22521–22543, doi:10.1029/2000JC000728, 2001.

Erickson, Z. K., A. F. Thompson, J. Callies, X. Yu, A. N. Garabato, and P. Klein,: The Vertical Structure of Open-Ocean Submesoscale Variability during a Full Seasonal Cycle. J. Phys. Oceanogr., 50, 145–160, https://doi.org/10.1175/JPO-D-19-0030.1, 2020.

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

Fu, L.-L., Pavelsky, T., Cretaux, J.-F., Morrow, R., Farrar, J. T., Vaze, P., et al.: The Surface Water and Ocean Topography Mission: A breakthrough in radar remote sensing of the ocean and land surface water. Geophysical Research Letters, 51, e2023GL107652. https://doi.org/10.1029/2023GL107652, 2024.

Gula, J., Taylor, J. R., Shcherbina, A., & Mahadevan, A.: Submesoscale processes and mixing. Ocean Mixing. DOI:10.1016/b978-0-12-821512-8.00015-3, 2022.

Haller, George.: Lagrangian Coherent Structures, Annual Review of Fluid Mechanics, DOI:10.1146/annurev-fluid-010313-141322, 2015.

Hou, M.; Yang, J.; Chen, G.: Eddy-Induced Chlorophyll Profile Characteristics and Underlying Dynamic Mechanisms in the South Pacific Ocean, Remote Sens, 16(14): 2628, https://doi.org/10.3390/rs16142628, 2024.

Hou, M. October 13, 2025 (v1)DatasetOpen A Global Sub-Mesoscale Eddy Dataset Using Deep Learning Algorithms and Sentinel-3 Chlorophyll-a Data (NetCDF) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.17483940, 2025.

Hou, M. A Global Sub-Mesoscale Eddy Dataset Using SWOT SLA data [Data set]. https://doi.org/10.5281/zenodo.20714411, 2026.

Hou, M., & Luo, N. Vortex identification result of 0.125° fused data [Data set], https://doi.org/10.5281/zenodo.20718880, 2026.

Konik, M., & Bradtke, K.: Object-oriented approach to oil spill detection using envisat asar images, Isprs Journal of Photogrammetry & Remote Sensing, 118, 37–52, https://doi.org/https://doi.org/10.1016/j.isprsjprs.2016.04.006, 2016.

Lévy, M., R. Ferrari, P. J. S. Franks, A. P. Martin, and P. Rivière.: Bringing physics to life at the submesoscale, Geophys. Res. Lett., 39, L14602, doi:10.1029/2012GL052756, 2012.

Lévy, M., Franks, P.J.S. & Smith, K.S.: The role of submesoscale currents in structuring marine ecosystems, Nat Commun 9, 4758, https://doi.org/10.1038/s41467-018-07059-3, 2018.

Li, R., Xia, L., Yang, J., Huang, B., & Chen, G.: Towards submesoscale eddy detection in SDGSAT-1 data through deep learning, International Journal of Digital Earth, 17(1), https://doi.org/10.1080/17538947.2024.2369625, 2024.

Liu, F., Zhou, H., Huang, W., Tian, Y., & Wen, B.: Cross-domain submesoscale eddy detection neural network for hf radar. Remote Sensing, 13 (13), 2441, https://doi.org/10.3390/rs13132441, 2021.

Liu, F., Tang, S., & Chen, C.: Satellite observations of the small-scale cyclonic eddies in the western south china sea. Biogeosciences, 12 (9), 299–305, https: //doi.org/10.5194/bg-12-299-2015, 2015.

Mason, E., A. Pascual, and J. C. McWilliams,: A New Sea Surface Height–Based Code for Oceanic Mesoscale Eddy Tracking. J. Atmos. Oceanic Technol., 31, 1181–1188, https://doi.org/10.1175/JTECH-D-14-00019.1, 2014

McWilliams, J. C.: Submesoscale currents in the ocean, Proc, 472 (2189): 20160117. https://doi.org/10.1098/rspa.2016.0117, 2016.

Ni, Q., Zhai, X., Wilson, C., Chen, C., & Chen, D.: Submesoscale eddies in the south china sea, Geophysical Research Letters, 48. https://doi.org/10.1029/2020GL091555, 2021.

Pan, H., Qiu, C., Liang, H., Zou, L., Zhang, Z., and He, B.: Different vertical heat transport induced by submesoscale motions in the shelf and open sea of the northwestern South China Sea, Front. Mar. Sci. 10:1236864. doi: 10.3389/fmars.2023.1236864, 2023.

Pegau, W. S., Boss, E., & Mart´ınez, A.: Ocean color observations of eddies during the summer in the gulf of california, Geophysical Research Letters, 29 (9). https://doi.org/10.1029/2001GL014076, 2002.

Reza, A. M.: Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology, 38(1), 35-44. https://doi.org/10.1023/B:VLSI.0000028532.53893.82, 2004.

Shcherbina, A. Y., E. A. D'Asaro, C. M. Lee, J. M. Klymak, M. J. Molemaker, and J. C. McWilliams,: Statistics of vertical vorticity, divergence, and strain in a developed submesoscale turbulence field, Geophys. Res. Lett., 40, 4706–4711, doi:10.1002/grl.50919, 2013.

Schubert, R., J. Gula, R. J. Greatbatch, B. Baschek, and A. Biastoch,: The Submesoscale Kinetic Energy Cascade: Mesoscale Absorption of Submesoscale Mixed Layer Eddies and Frontal Downscale Fluxes, J. Phys. Oceanogr., 50, 2573–2589, https://doi.org/10.1175/JPO-D-19-0311.1, 2020.

Sukigara, C., Inoue, R., Sato, K., Mino, Y., Nagai, T., Fassbender, A. J., Takeshita, Y., Bishop, S., and Oka, E.: Observing intermittent biological productivity and vertical carbon transports during the spring transition with BGC Argo floats in the western North Pacific, Biogeosciences Discuss. [preprint], https://doi.org/10.5194/bg-2022-9, 2022.

Suzuki, S., & Be, K.: Topological structural analysis of digitized binary images by border following. Computer Vision Graphics & Image Processing, 30(1), 32-46, DOI:10.1016/0734-189X(85)90016-7, 1985.

Verger-Miralles, E., Mourre, B., Gómez-Navarro, L., Barceló-Llull, B., Casas, B., Cutolo, E., et al.: SWOT enhances small-scale eddy detection in the Mediterranean Sea, Geophysical Research Letters, 52, e2025GL116480, https://doi.org/10.1029/2025GL116480, 2025.

Vidhya, G. R., & Ramesh, H.: Effectiveness of contrast limited adaptive histogram equalization technique on multispectral satellite imagery, In Proceedings of the International Conference on Video and Image Processing (pp. 234-239), https://doi.org/10.1145/3177404.3177409, 2017.

Williams, S., Hecht, M., Petersen, M., Strelitz, R., Maltrud, M., Ahrens, J., Hlawitschka, M. and Hamann, B.: Visualization and Analysis of Eddies in a Global Ocean Simulation. Computer Graphics Forum, 30: 991-1000. https://doi.org/10.1111/j.1467-8659.2011.01948.x, 2011.

Xia, L., Chen, G., Chen, X., Ge, L., & Huang, B.: Submesoscale oceanic eddy detection in SAR images using context and edge association network. Front. Mar. Sci. 9:1023624. doi: 10.3389/fmars.2022.1023624, 2022.

M. Xu, H. Li, Y. Yun, F. Yang and C. Li,: End-to-End Pixel-Wisely Detection of Oceanic Eddy on SAR Images With Stacked Attention Network, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 9711-9724, 2023, doi: 10.1109/JSTARS.2023.3322404, 2023.

https://doi.org/10.5194/essd-2026-475
Preprint. Discussion started: 13 July 2026

Roemmich, D., G.C. Johnson, S. Riser, R. Davis, J. Gilson, W.B. Owens, S.L. Garzoli, C. Schmid, and M. Ignaszewski.: The Argo Program: Observing the global ocean with profiling floats, Oceanography 22(2):34–43, https://doi.org/10.5670/oceanog.2009.36, 2009.

Yamada, K., Ishizaka, J., Yoo, S., Kim, H. C. , & Chiba, S.: Seasonal and interannual variability of sea surface chlorophyll a concentration in the Japan/East Sea (JES), Progress in oceanography, 61(2-4), 193-211, https://doi.org/10.1016/j.pocean.2004.06.001, 2004.

Y. Liu, X. Li and Y. Ren,: A Deep Learning Model for Oceanic Mesoscale Eddy Detection Based on Multi-source Remote Sensing Imagery, IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 2020, pp. 6762-6765, doi: 10.1109/IGARSS39084.2020.9323716, 2020.

Zhang, Z., Tian, J., Qiu, B. et al.: Observed 3D Structure, Generation, and Dissipation of Oceanic Mesoscale Eddies in the South China Sea, Sci Rep 6, 24349, https://doi.org/10.1038/srep24349, 2016.