Statistical Features of Spatiotemporal Distributions of Internal Waves in the Northern South China Sea Derived from MODIS Imagery by Deep Learning Approach

Xudong Zhang[1, 2], Xiaofeng Li[1, 2*]

[1] Key Laboratory of Ocean Observation and Forecasting
[2] Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China

Correspondence to: Xiaofeng Li (lixf@qdio.ac.cn)

Abstract Internal waves (IWs) are an important ocean process in transmitting energy between multiscale ocean dynamics, making them a crucial oceanic phenomenon. The South China Sea (SCS) is renowned for its frequent large-amplitude IW activities, emphasizing the importance of collecting and analyzing extensive observational data. In this study, we present a comprehensive IW dataset covering the northern SCS covering the area from 112.40°E to 121.32°E and from 18.32°N to 23.19°N, spanning from 2000 to 2022 with a 250 m spatial resolution. The IW dataset comprises 3085 high-resolution MODIS true-color IW images paired with precise IW position information extracted from 15830 MODIS images using advanced deep learning techniques (DOI: 10.12157/IOCAS.20240409.001). IWs in the northern SCS are divided into four regions based on extracted IW spatial distributions, facilitating detailed analyses of IW characteristics, including spatial and temporal distributions across both the entire northern SCS and its sub-regions. Notably, we uncover typical "double-peak" distributions corresponding to the lunar day, underscoring IWs' close relationship with tides. Furthermore, we identify two IW-free silence regions attributed to underwater topography influences, indicating varied IW characteristics across regions and suggesting underlying mechanisms warrant further investigation. The constructed dataset holds significant potential for applications in studying IW-environment interactions, developing monitoring and prediction models, validating and enhancing numerical simulations, and serving as an educational resource to foster awareness and interest in IW research.

1 Introduction

Oceanic internal waves (IWs) are a featured phenomenon in marginal seas and continental shelf ocean areas, characterized by their horizontal propagation over long distances and large amplitude within stratified water (Haury et al., 1979; Magalhaes et al., 2020; Magalhaes et al., 2022; Pan et al., 2007; Zhang et al., 2022; Zhao et al., 2014). Their significance lies in their role as transmitting energy between multiscale ocean dynamics and their critical impact on the ocean environment, ocean acoustics, and underwater navigation (Jia et al., 2019; Ramp et al., 2022b). Internal waves appear either as a periodic series of waves distinguished by their amplitude and crest length or as solitary solitons. While the IW crest length extends several hundreds of kilometers, the characteristic length of IWs along its propagation direction extends only several hundred meters to a few kilometers. The isolation feature and infrequent occurrence render these waves challenging to capture. However,

comprehending IWs needs extensive collection and analysis of observational data. Traditional methods using oceanic exploration instruments are costly, labor-intensive, and unsuitable for large-scale observations due to the submerged nature of IWs.

35   A viable solution to this challenge is offered by remote sensing techniques benefitting from its rapid response, large spatial coverage, and cost-efficiency (Li et al., 2008; Zhang et al., 2019). Over the past 20 years, the amount of satellite data has grown exponentially which has enabled the construction of an IW dataset at a larger and longer scale. Leveraging the ocean surface convergent and divergent motions induced by IWs, satellite-based IW imaging relies on identifying surface patterns of calm and roughened waters (Zheng et al., 2001). Sea surface roughness is a predominant factor affecting the backscattering

40   intensity of active microwave sensors, such as the synthetic aperture radar (SAR) (Furtney et al., 2024; Jia et al., 2018; Zhao et al., 2004). However, passive sensors, such as radiometers, can detect the IW-induced sea surface roughness signatures by receiving sunlight reflected by the ocean surface (De Macedo et al., 2023; Hu et al., 2021; Sun et al., 2021). For instance, since 2000, data with nearly daily global monitoring at a spatial resolution of 250 m has been provided by the Moderate-resolution Imaging Spectroradiometer (MODIS) on the Aqua and Terra satellites, suitable for more in-depth IW investigation since it

45   achieves the best possible balance between orbital duration and spatial coverage (De Macedo et al., 2023).

The advent of cloud computing platforms, like the Earth Observation (EO) Browser from ESA, the WorldView from NASA, and the Google Earth Engine (GEE) from Google, have liberated us from the repetitive and arduous image pre-processing steps (e.g., radiometric, atmospheric, and geometric corrections). Therefore, the foremost challenge encountered during dataset

50   construction is accurately detecting and obtaining the limited IW information concealed in the massive satellite observations. While extracting the IW crest manually can reduce errors, it also increases processing time. The primary foundations of conventional automatic or semi-automated extraction approaches utilizing fundamental image processing techniques are image segmentation and edge detection (Kurekin et al., 2020). Nonetheless, the edge detection algorithm often results in discontinuous edge pixels that may not characterize a complete IW crest. On the other hand, image segmentation techniques

55   fail to determine consistent threshold values and need additional processing steps to detect boundary pixels.

Deep convolutional neural networks (DCNNs) have showcased their capacity in image pattern classification and have become a dependable tool for extracting accurate pixel-level targets from oceanic remote sensing imagery (Li et al., 2022; Li et al., 2020; Liu et al., 2019; Wang and Li, 2023). Numerous machine-learning techniques have been put forth for the automatic

60   extraction of the IW crest from geostationary optical and space-borne SAR imagery (Bao et al., 2019; Ma et al., 2023; Tao et al., 2022; Zheng et al., 2021). Nevertheless, these studies solely conducted experimentation and validation on individual sensors and restricted geographical regions with few images, which are inadequate for producing a comprehensive IW database over an extended temporal period. Recently, Zhang et al. (2023) developed a robust DCNN-based IWE-Net (IW extraction

2

network) model for automatically extracting IW signatures from several satellite sensors with different spatial resolutions, even in difficult imaging circumstances.

The northern South China Sea (SCS) serves as an exceptional natural laboratory for studying IWs of large amplitude (Alford et al., 2015; Bai et al., 2017; Bai et al., 2014; Cai et al., 2012; Guo and Chen, 2014; Liang et al., 2019; Liu and Hsu, 2004; Ramp et al., 2022a). IW propagation characteristics, such as the reflection, refraction, and shoaling process have been extensively studied by scholars. Besides active IW activity, the northern SCS also has circulation patterns, eddies, Kuroshio intrusion, and other dynamic processes, which may influence the IW features (Dong et al., 2016; Liu et al., 2014; Liu and Abernathey, 2023; Liu et al., 2022; Liu et al., 2016; Xu et al., 2020). Considering the multi-scale dynamic process and active IW activity, a long-time series IW dataset could benefit the study of multi-scale dynamic process interactions. In this study, we initially employ the IWE-Net to extract IWs from the entire set of MODIS images acquired over 22 years in the northern SCS. After essential post-processing steps, we establish an accessible and extensive IW dataset to enhance the availability of resources for pertaining research, such as different life stages of IWs.

The paper is organized as follows: section 2 describes the satellite images and the deep-learning model; Section 3 presents the results; Section 4 presents the new findings from the built dataset; and Section 5 shows the conclusion and outlook of the dataset.

2 Data and Methods

2.1 MODIS Imagery Collection

MODIS sensors are situated at a height of roughly 700 kilometers in sun-synchronous orbits. The National Aeronautics and Space Agency (NASA) launched Terra and Aqua, two Earth Observation System (EOS) satellites, in December 1999 and May 2002, respectively. With a range of spatial resolutions from 250 m to 1 km (bands 1 and 2 are 250 m resolution, bands 3–7 are 500 m resolution, and bands 8–36 are 1 km resolution), the satellite constellation provides almost daily coverage of the whole Earth by gathering imagery over a 2300-km wide swath. Pre-processing MODIS data typically involves several steps, such as downloading the original hierarchical data format (HDF) files, geographical correction, radiometric calibration, atmospheric correction, re-projection, etc. These intricate procedures act as necessities but impede the analysis of long-term and large-scale ocean phenomena, falling short of meeting the demands of dataset construction.
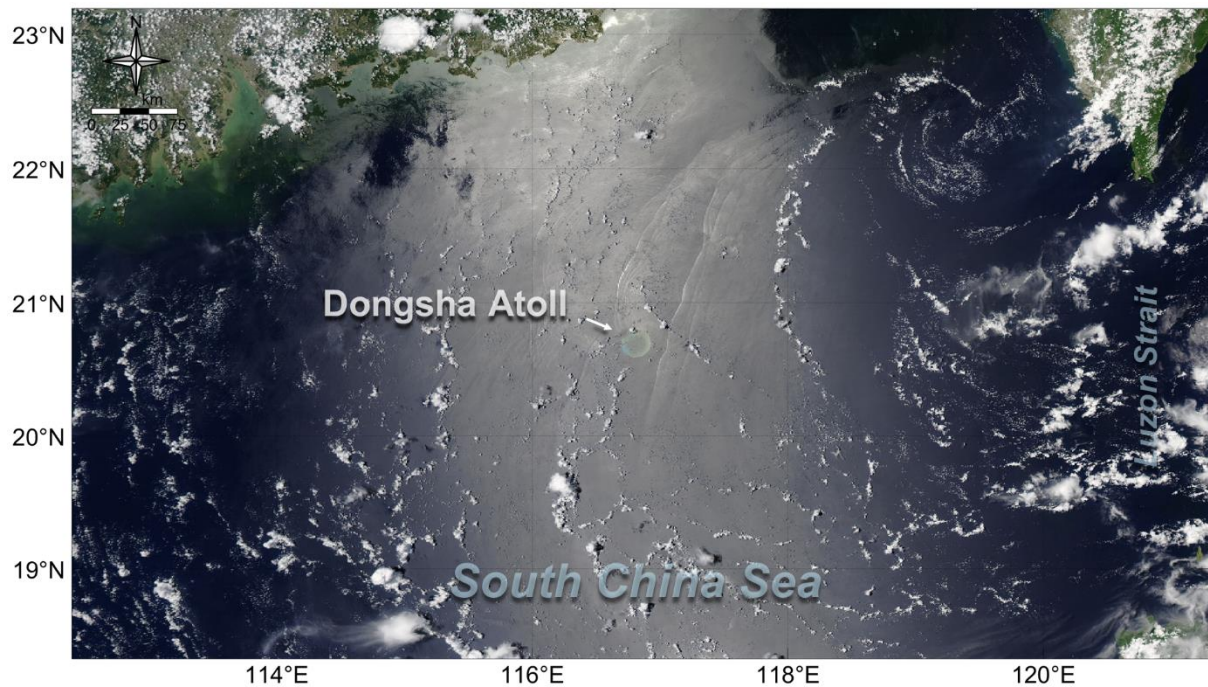
Figure 1. An example of Worldview's natural color MODIS image in the northern South China Sea with IW signatures at the Dongsha Atoll was acquired on Jul. 14, 2021, by Aqua.

95　Through an interactive browsing experience, users can explore global and full-resolution satellite images stored by the Global Image Browse Services (GIBS) system using NASA's Worldview (https://worldview.earthdata.nasa.gov/). The MODIS Corrected Reflectance products (Figure 1) leverage Level 1B data (the calibrated, geolocated radiances) to provide natural color images (the red channel derived from band 1, the green channel from band 4, and the blue channel from band 3). This process also involves the removal of significant atmospheric effects, including Rayleigh scattering, to enhance the image

100　quality. Worldview offers Terra MODIS products from Feb. 25, 2000, and Aqua products from Jul. 04, 2002. The target area covers 112.40-121.32°E and 18.32-23.19°N. We collected 15830 MODIS natural color images from 2000 to 2022 as model input, with 8345 from Terra and 7485 from Aqua. All these images have a 250 m spatial resolution and are stored in a GeoTIFF format, which embeds geospatial information into image files.

2.2 Deep Learning Model

105　The goal of the deep-learning model IWE-Net (Zhang et al., 2023) is to extract IW locations from numerous satellite images obtained from optical and SAR sensors operating in sun-synchronous or geostationary orbits with different spatial resolutions. This model underwent training and testing using a dataset comprising 1115 satellite images, encompassing 116 full-swath Environmental Satellite (ENVISAT) Advanced SAR (ASAR) images, 839 Terra/Aqua MODIS images, and 160 Himawari-8 Advanced Himawari Imager (AHI) images. All these satellite images have clear IW signatures in the SCS, Sulu Sea, and

110　Celebes Sea. Three major improvements are incorporated into IWE-Net to increase its resilience and accuracy: squeeze and

excitation blocks, online data augmentation, and the Matthews correlation coefficient loss function, which takes into consideration the distinct properties of IW under various imaging techniques. The structure of the IWE-Net is presented in Figure 2.
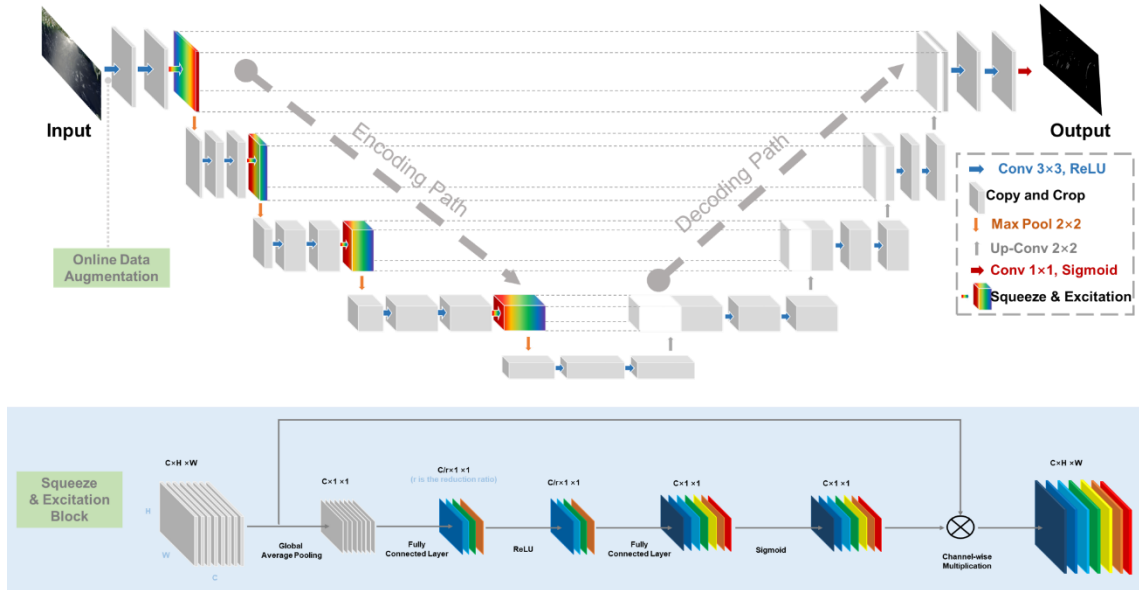
Figure 2. IWE-Net model structure with three tailored modifications adapted from Zhang et al. (2023).

We employ the Pixel Accuracy, Precision, Recall, and F1-score as metrics to evaluate the positional differences between the IW dataset and the ground truths. Pixel Accuracy represents the proportion of the image's pixels that were properly classified. When there is a significant percentage of negative samples (non-IWs), such as in this task, the Pixel Accuracy often approaches 1 and exhibits a limited responsiveness. When managing uneven classes, Precision, Recall, and F1-score are suitable metrics to evaluate the classifier's output quality. Precision reflects the proportion of the false IW pixels in the dataset, while Recall indicates the proportion of the missed ones. F1-score is the harmonic mean of these two metrics, offering a balance between Precision and Recall. The testing set boasts an overall mean precision of 85.75%, recall of 85.67%, and F1-score of 85.71%, demonstrating the model's accuracy in extracting IW signatures.

2.3 Post-processing

IWE-Net's performance in the SCS using MODIS images exhibits a mean Precision of 87.90%, which implies that there are approximately 12% false classifications within the model's results caused by a small minority of IW-like features, such as aircraft trails, linear and sparse clouds, as well as surface signals of shallow water topography and plumes. These small-scale misclassifications, characterized by their varying shapes and orientations but consistent positions, can be readily eliminated using manual methods, thus contributing to an overall improvement in the accuracy of this IW dataset. Since the model-produced IW locations are stored in longitude and latitude, users can do more post-processing procedures as they need.

2.5 Data Records

This study generated two sets of data: true color MODIS Imagery with IW occurring and the positional information of the IWs. All data has been archived and is stored in the XXXXXX repository.

- MODIS IW imagery
  - Repository Location: https://XXXXXXXXXXXXX
  - Data Format: GeoTIFF. The format stands out as an advantageous choice for storing MODIS imagery of IWs, primarily due to its capability for embedding georeferencing information (in the World Geodetic System 84) directly into the TIFF (Tag Image File Format) file, ensuring precise mapping of pixels to geographical locations. Given its broad-based endorsement, GeoTIFF guarantees interoperability across various Geographic Information System (GIS) platforms and image manipulation utilities, concurrently providing robust support for extensive datasets.
  - File Structure:
    - Naming Convention: MODIS_TrueColor_YYYY-MM-DD_SSS.tiff (where YYYY-MM-DD represents the acquisition date of the image and SSS represents the satellite Terra or Aqua)
    - Image Size: 4061 (width) x 2218 (length) pixels
    - Resolution: A 250 m x 250 m region on the ground is represented by each pixel.
    - Data Layers: The following data layers are included:
      - Red channel (Band 1): Data range [0, 255]
      - Green channel (Band 4): Data range [0, 255]
      - Blue channel (Band 3): Data range [0, 255]
    - Georeferencing Information (in the Metadata): Includes projection system, image size, resolution, etc.
- IW Position Information
  - Repository Location: https://XXXXXXXXXXXX
  - Data Format: Shapefile. This format distinguishes itself in the archival of IW position data, attributable to its pervasive compatibility with diverse GIS software, thereby assuring seamless interoperability and efficacious data dissemination. It permits the execution of sophisticated spatial analyses directly upon the data, engendering a nuanced understanding of IW phenomena. The robust framework of this format encapsulates both geometric and attributive information, upholding the integrity of the data, while its proficiency in handling voluminous datasets guarantees expeditious and reliable access to pertinent information.
  - File Structure:
    - Naming Convention: IW_YYYY-MM-DD.shp (where YYYY-MM-DD represents the date the IWs occurred)
    - Column Names and Data Types:
      - longitude: Float, precision to 4 decimal places
      - latitude: Float, precision to 4 decimal places

3 Signature Extraction and Validation

IWE-Net is designed as an end-to-end model, signifying that both the input and output of the model are images. The task of IW location extraction is framed as a binary classification problem, where the output image exclusively contains two values: 1 denoting the presence of IWs and 0 representing non-IW features. The natural color product's red channel (derived from Band 1) was selected as the input of IWE-Net. Figure 3 illustrates an example of the output and the corresponding input image acquired on Aug. 28, 2002. The extraction results show that most of the IWs were located around the Dongsha Atoll, like the distribution in previous studies. However, it is noteworthy that IWE-Net can successfully identify IWs even in darker regions,

as exemplified in the southwest portion of Figure 3, which is distantly located away from the sun glint area and can hardly be seen with the naked eye if no image enhancement is performed. It suggests that deep-learning-based extraction models have the potential to generate results that could surpass the accuracy achieved through visual interpretation, particularly when

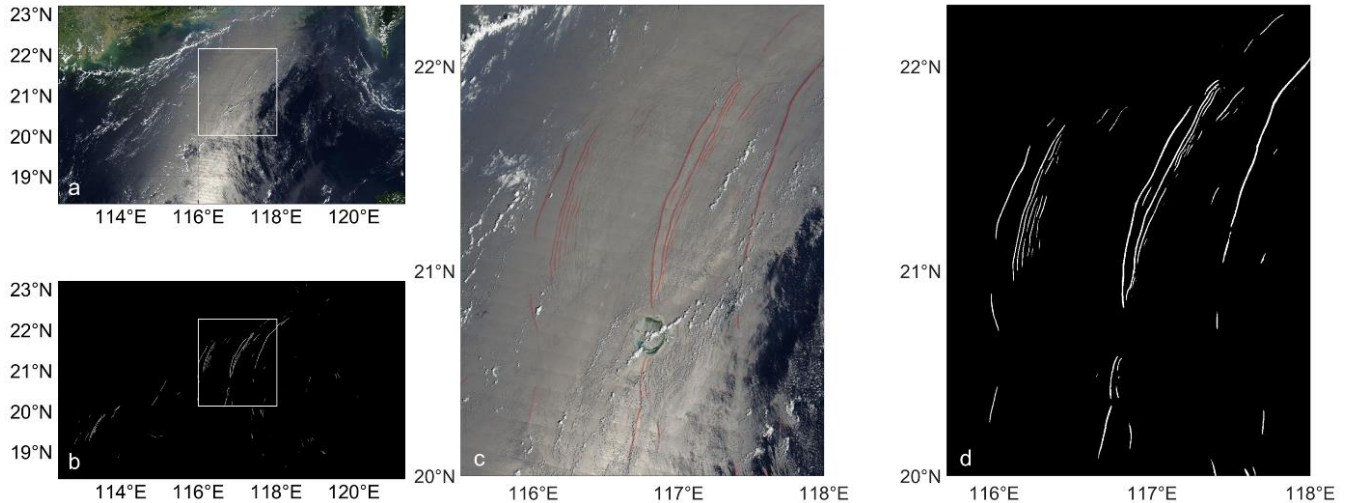175 handling extensive datasets. Altogether, 3085 MODIS images containing IW signatures were identified among the 15830 input images.

Figure 3. An example of IWE-Net's output (b) with the MODIS image (a) was acquired on Aug. 28, 2002. Panels (c) and (d) are enlarged regions denoted by the white box in panels (a) and (b). The red lines in panel (c) correspond to the white lines in panel (d).

180

As shown in Figure 3, the white points within the image matrix represent values predicted as 1 by the model, while the remaining points are all predicted as 0 (rendered in black). Consequently, evaluating the model's extraction accuracy hinges on IWs' positional extraction precision. Nonetheless, no corresponding IW products are available in the SCS, and the precision of manual extraction surpasses that of other current approaches due to the complicated imaging conditions for MODIS images

185 in the SCS. Ground-truth maps derived from labels provided with visual interpretation were used to assess the overall accuracy of this IW crest dataset. In practical implementation, we initiated the process by introducing a new layer atop the MODIS image to maintain consistency in the referenced IW image size. Subsequently, we carefully indicate the IW spots using white lines and then give the layer a black background.
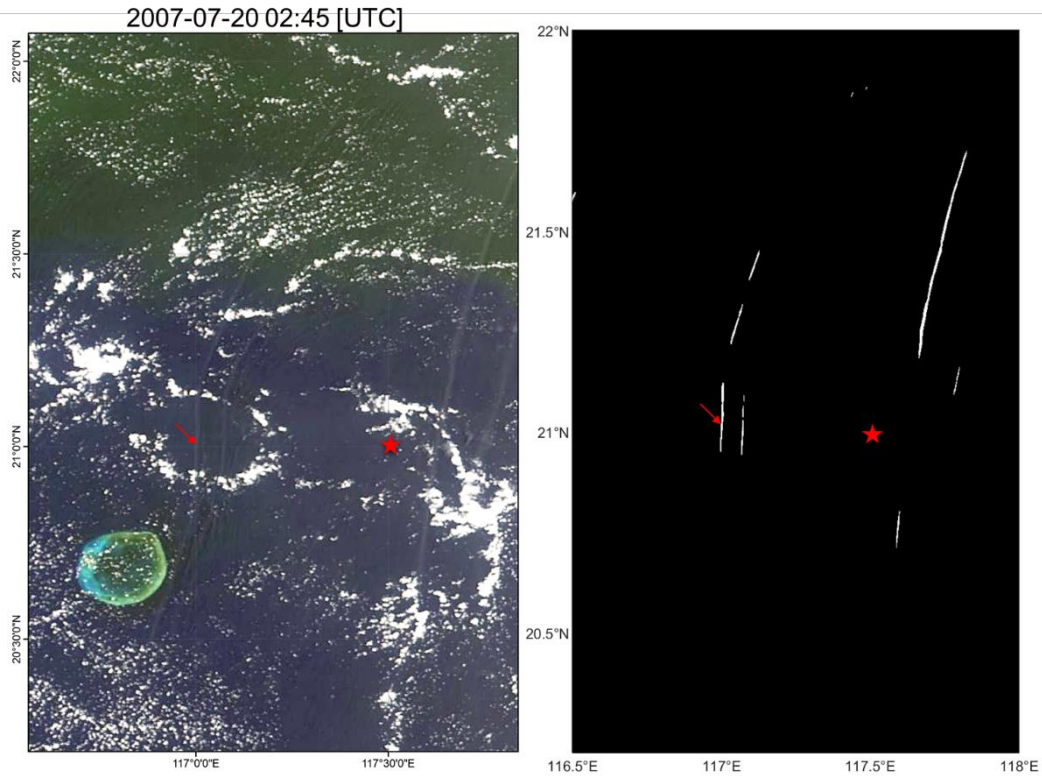
Figure 4 illustrates an example of IW detection using a MODIS image captured on 20 July 2007 at 02:45 UTC, alongside field observations detailed by Zhao et al. (2012). The red star marks the locations of the field observations, while the red arrow indicates the IW observed in these field studies. According to Zhao et al. (2012), the IW had an amplitude of several tens of meters and vertical wave-induced currents exceeding 0.5 m/s (see Figure 3 in their work). This IW was effectively detected through field observations and subsequently captured by the Terra MODIS image approximately 7 hours later. The near-synchronous detection of IWs from both satellite imagery and field observations provides strong validation for the accuracy of the applied model and produced dataset.

4 Statistical Analysis

4.1 ISW Spatial Distributions in the Northern SCS

We superimpose the IWE-Net-produced IW crest lines using MODIS images from 2000 to 2022 and the result is shown in Figure 5. The spatial resolution of the superimposed map is 250 m, which is the same as the input MODIS image. We can find that most IWs are located around the Dongsha Atoll, and four IW clusters are located in deep and shallow ocean areas.

Generally, more IWs are found in the continental shelf regions than in the deep ocean region. The IW distributions agree well with the topography feature.

As shown in Figure 5, we divided the detected IW locations into four regions 1-4, which cover the area from 112.5°E to 114.2°E and from 18.5°N to 20.9°N, from 114.2°E to 118.1°E and from 19.5°N to 22.2°N, from 118.1°E to 120.0°E and from 22.0°N to 23.0°N and from 118.1°E to 120.5°E and from 19.5°N to 22.0°N, respectively. The division was based on the geometry of IW crests, which suggest distinct sources for Regions 1 and 3, and different life stages of IWs before and after IWs propagate from the deep ocean to the continental shelf areas in Regions 2 and 4. IWs in Regions 1, 2, and 4 mainly propagate westward, while IWs in Region 3 propagate southward, which implies different IW generation sources. More IWs are observed in Region 2 because IWs in Region 4 are mainly solitons. When these IW solitons propagate into shallower regions in Region 2, where water depth is less than 1000 m, IW soliton will fission into IW packets. In addition, the existence of the Dongsha Atoll will cause IW reflection or refraction, which makes the IW characteristic more complicated (Jia et al., 2018; Li et al., 2013). The IW wave crests in Region 1 are not always aligned with IWs in Region 2 and Region 4, which also promise different IW generation sources or mechanisms.
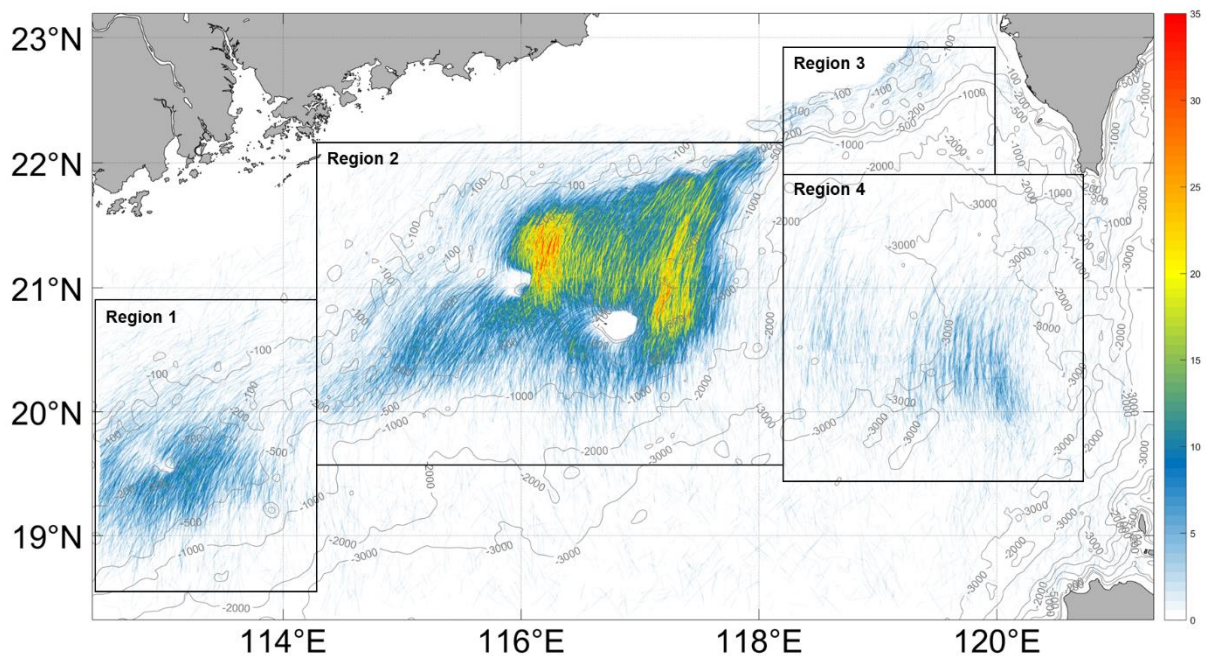
Figure 5. Superimpose the IW detection results from MODIS images from 2000 to 2022. The color indicates the frequency of IW observed at each location. The map resolution is 250 m which is the same as the input MODIS image.
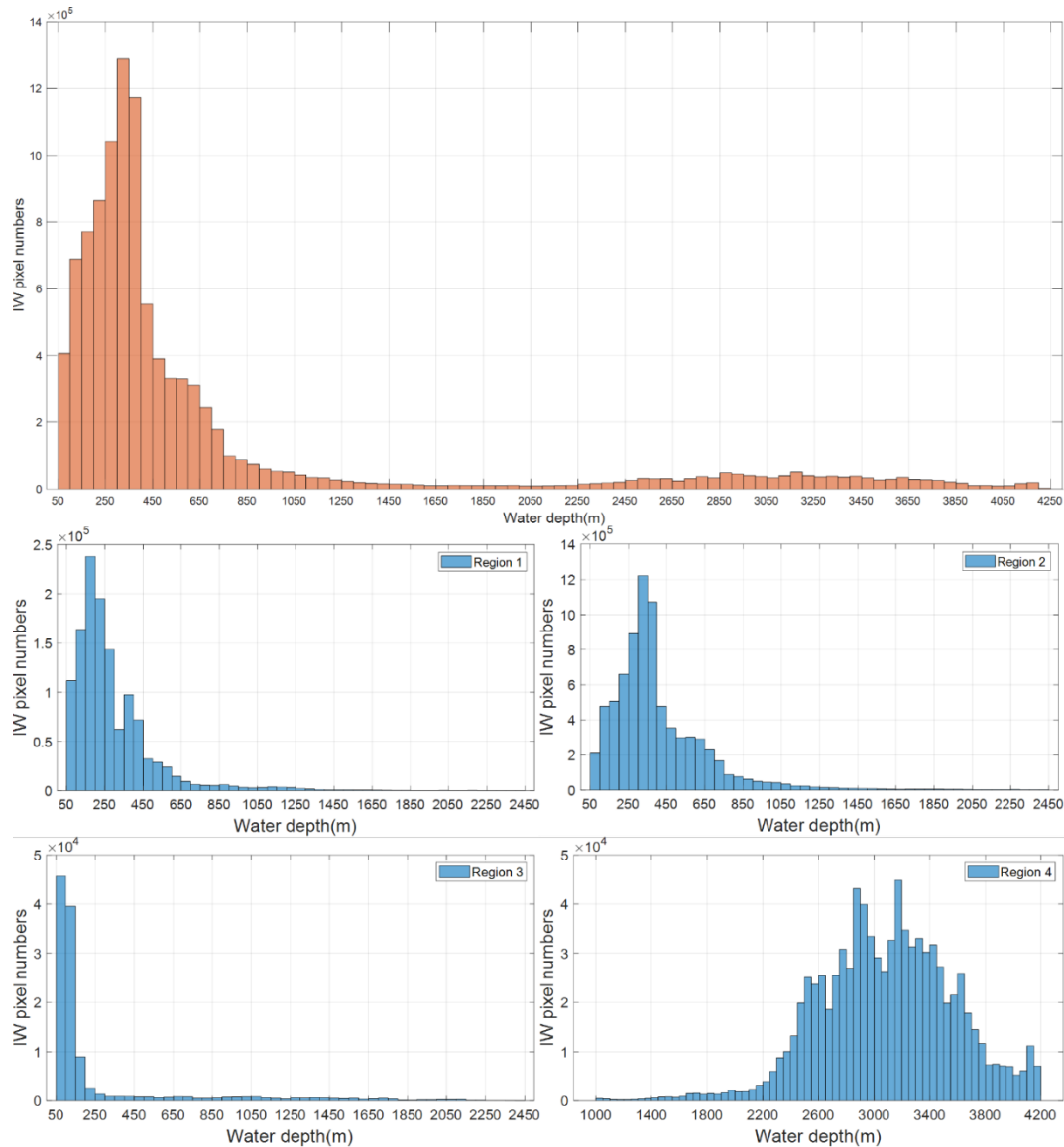
9

Figure 6. Statistical histograms of IW pixel number vs the water depth for (a) the whole northern SCS (orange) and (b-e) four regions as shown in Figure 4 (blue)

Underwater topography significantly influences internal wave (IW) evolution. Figure 6 illustrates IW distribution relative to water depth, indicating a prevalence of IWs in open ocean areas with depths under 1000 m. Interestingly, there are more IWs at depths of 3000 m than 2000 m, hinting at an IW evolution mechanism that warrants further study. In Regions 2 and 4, as shown in Figure 4, the distribution of IW clusters reflects specific depth characteristics. Specifically, Region 1 features IWs mainly in depths less than 600 m despite the overall range from 100 m to 2000 m. For Region 2, IWs predominantly occur at depths under 1000m, aligning closely with water depth contours between 100 m and 1000 m. In Region 3, IWs are primarily

10

found at 100 m depths, moving away from the continental shelf and vanishing beyond 2000m. Region 4's IWs are mostly at 2600 to 3600 m depths, rarely below 2000 m, showcasing a general correlation between IW distribution and water depth contours.

4.2 ISW Temporal Distributions-Monthly Variations

Stratification stands out as its pivotal influence on the generation and propagation dynamics of IWs. Stratification shows significant seasonal variations in the northern SCS and thus results in IW distribution variation. Figure 7 illustrates the monthly variations of IWs detected by the IWE-Net. One can find that IW occurrences are prevalent from May to August, whereas sightings diminish notably in other months. This temporal disparity underscores the influence of seasonal changes on the stratification and IW activity. Shallower depths and intensified stratification during these warmer months favor heightened IW activity, facilitating their generation and propagation across the northern SCS. Conversely, stratification deepens while weakening during winter, attributed to factors such as intensified winter monsoons. The resultant stratification dynamics during this period are less conducive to IW generation and propagation, contributing to fewer observed IW events. Moreover, the weakened modulation of surface features by IWs during winter further diminishes their detectability. The classified four regions exhibit a similar trend to the entire northern SCS. However, it's noteworthy that in Region 3, the distribution is more concentrated in July, suggesting that IWs in this area may require more stringent conditions for generation. These findings underscore the seasonal modulation of stratification and its consequential impact on IW dynamics. They highlight the intricate interplay between atmospheric phenomena, such as monsoonal circulations and solar radiation, in shaping the observed seasonal variations in IW activity within the northern SCS.
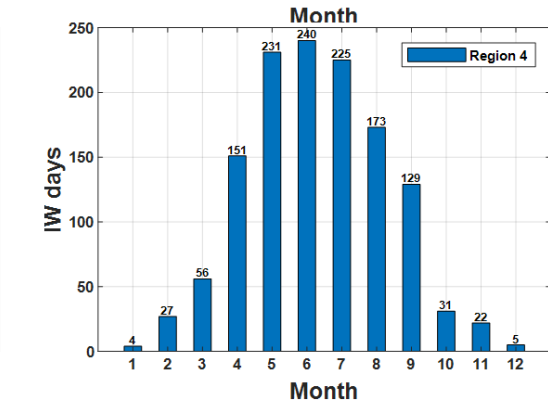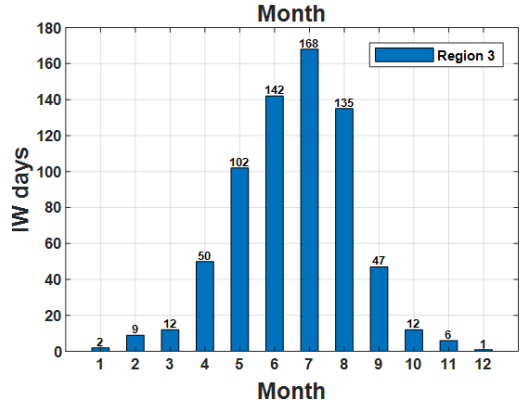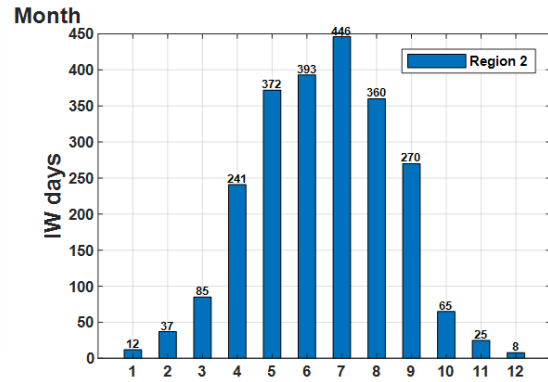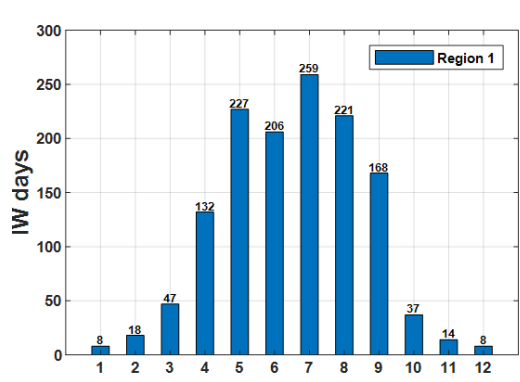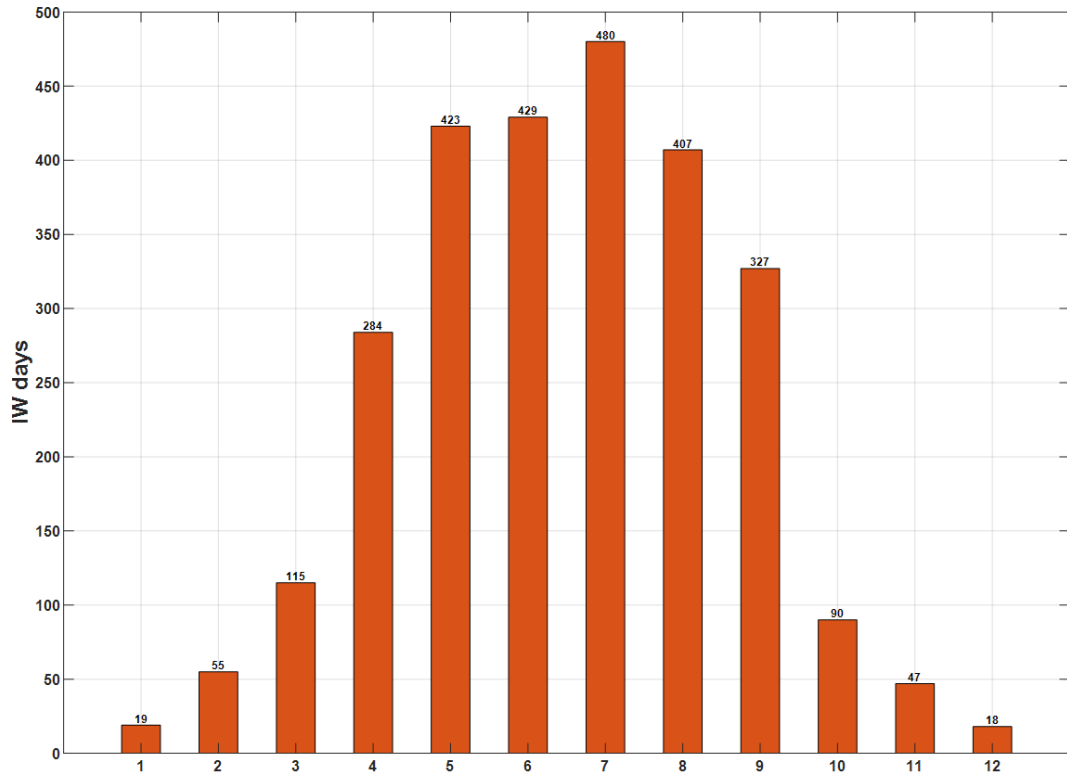
Figure 7. Statistical histograms of monthly distributions of IW days for (a) the whole northern SCS (orange) and (b-e) four regions (blue).

4.3 IW Variations with the Lunar Day

255   The generation of IW exhibits a close correlation with the astronomical tide. The magnitude of the tide directly impacts the scale of IWs. Astronomical tides follow a fortnightly pattern, with the first and fifteenth lunar days of the month coinciding with the period of spring tides. Consequently, IW characteristics display fortnightly variations. Illustrated in Figure 8 are the IW variations relative to the lunar day. It is evident that IWs in the northern SCS typically exhibit a "double-peak" distribution pattern. The peak IW occurrences occur approximately four days after the spring tide. IWs in this region require roughly four
260   days to propagate from the Luzon Strait to the continental shelf regions before eventually dissipating.

All four regions display a "double-peak" pattern linked to tidal dynamics. Regions 2 and 4 are the main channels for IW in the northern SCS and record the highest number of IW observation days. Region 3, in contrast, has the lowest, with 35 days marking the highest count in this region, indicating that IW generation and propagation conditions in Region 3 are more
265   stringent and less favorable than in Region 2, which has a minimum of 35 observation days. Region 4 shows more distinct and focused peaks than Region 2 due to the dominance of IW solitons in Region 4, while Region 2 primarily exhibits IW packet behavior. The shallower water in Region 2 results in a slower IW propagation speed than in Region 4, leading to broader peaks in Region 2.
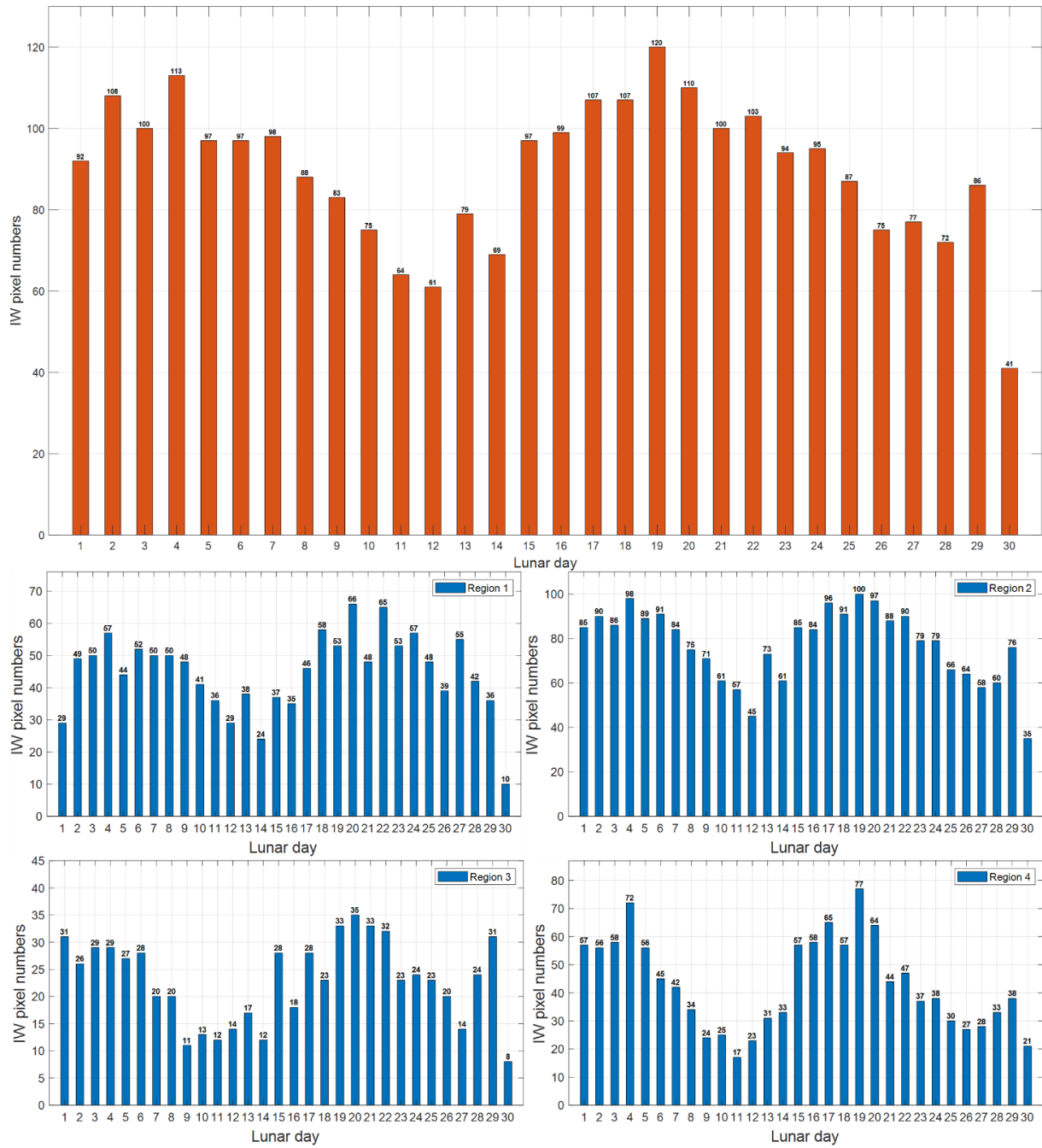
Figure 8. Statistical histograms of IW pixel number vs the lunar day for (a) the whole northern SCS (orange) and (b-e) four regions (blue)

4.4 IW Silence Regions

Figure 5 shows three distinct blank areas in Regions 1 and 2. One such area in Region 2 encompasses the Dongsha Atoll, while the other undistinguished area behind the Dongsha Atoll has received minimal attention in prior studies. These blank spaces
275 signify limited or absent IW activity, delineating what we refer to as IW-free silence regions. Figure 9 reveals the presence of a chain of small underwater ridges situated in the northwest direction of the Dongsha Atoll. These ridges correspond to a series of IW-free silence regions, as indicated by two black arrows. The unique underwater topography contributes to forming IW-free areas within the northern SCS.

280 The IW-free region adjacent to the Dongsha Atoll extends approximately 110 km towards the continental shelf area, with less than 100 m water depths. Conversely, the IW-free region in Region 1 is comparatively smaller, characterized by conspicuous underwater ridges aligned along the direction of IW propagation. These underwater ridges segregate IW crests, with subsequent reconnection occurring at 112.7°E.
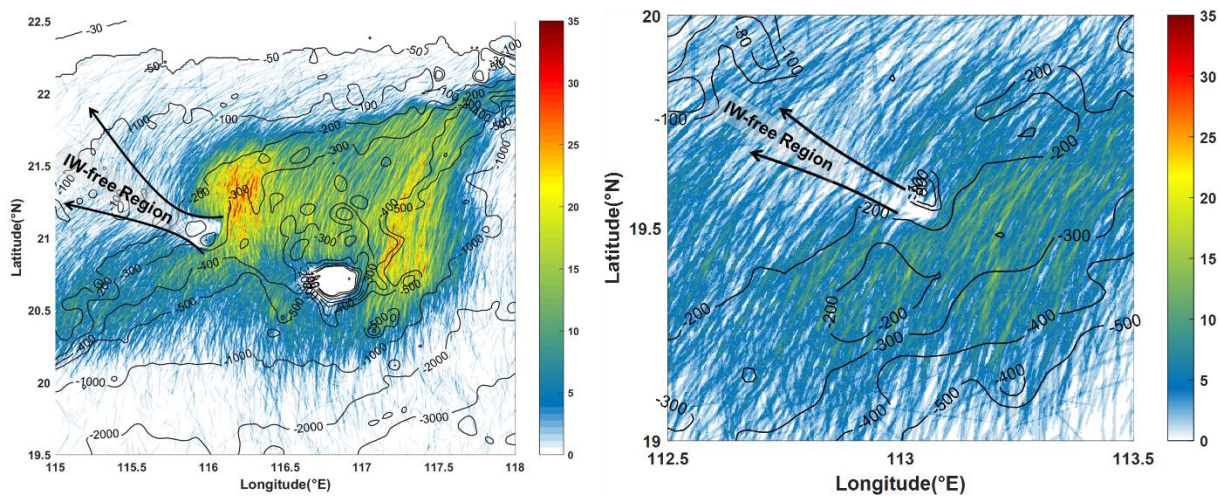

285 **Figure 9. IW silence regions (black arrows) in (a) Region 2 and Region 1 derived from superimposed IW location**

IWs exhibit a widespread distribution across the northern SCS, yet noticeable gaps between distinct IW wave crest clusters are evident. Illustrated in Figure 10 are two such IW gaps observed in Region 4, with spatial separations of 63.6 km and 31.7 km. This occurrence can be attributed to the predominantly solitonic nature of IW in Region 4, characterized by phase speeds exceeding 3.0 m/s. The rapid propagation of IWs, coupled with two MODIS snapshots captured each day, likely contributes
290 to the formation of these gaps.

Another IW gap, spanning a spatial distance of 62.6 km, is observed between Regions 1 and 2. The disparate directions of IW wave crests in the area suggest distinct IW generation sources. As a result, IWs originating from different sources undergo separate evolution processes and fail to connect. Additionally, IWs in Region 2 exhibit a gap with Region 1 at the southern

15

295   edge and another gap with Region 3 at the northern edge. Notably, the gap between Region 2 and Region 3 coincides with abrupt underwater topography and the presence of small underwater ridges.
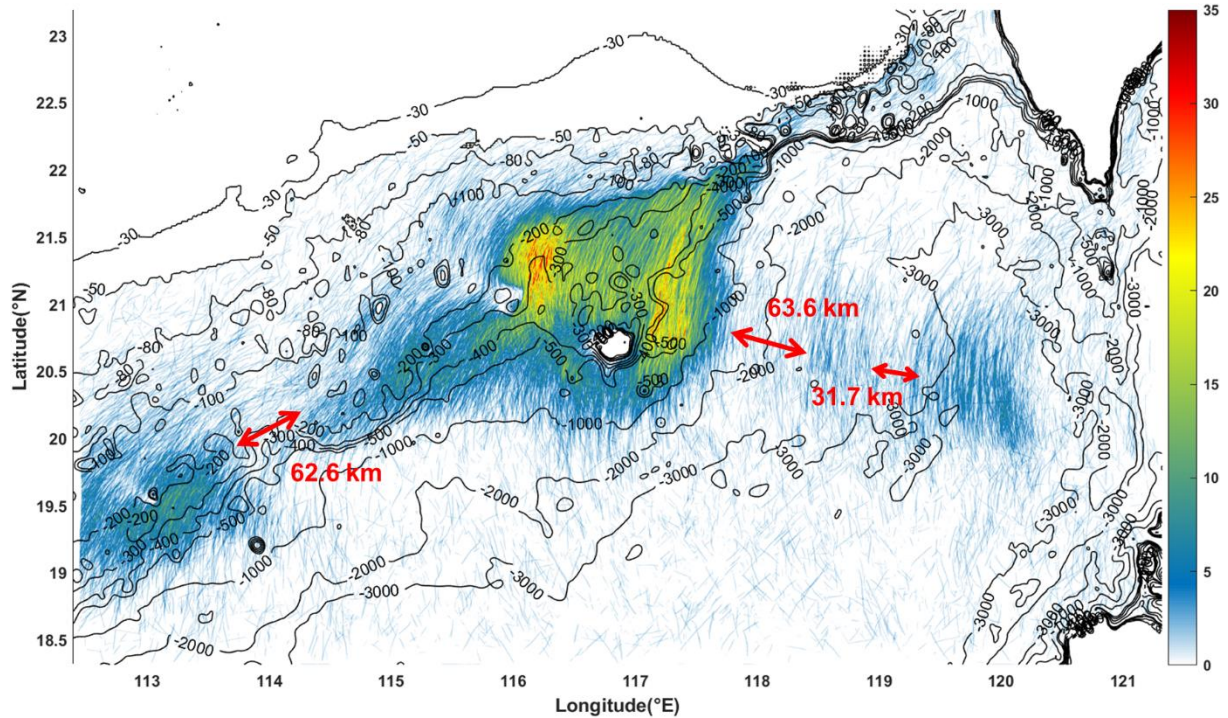

Figure 10. Gaps between different IW clusters in the northern South China Sea.

5. Conclusion and Outlooks

300   In this study, we have successfully constructed a comprehensive oceanic IW dataset from 2000 to 2022, utilizing MODIS satellite imagery and the IWE-Net model. Through applying advanced remote sensing images and deep learning algorithms, we have extracted precise IW locations, specifying longitude and latitude coordinates, and organized them into a Shapefile format for easy access and analysis.

305   Meanwhile, it is important to note that the main sources of error in the constructed IW dataset are twofold: on one hand, while optical imagery can capture most IW features, weather conditions such as clouds and rain may prevent MODIS imagery from detecting IWs even if they are present. On the other hand, polar orbiting satellites may have uneven coverage in certain areas and gaps between orbits, which can lead to some missed IW detections in the model's results. In the future, considering including more satellite sensors for detecting IWs, especially SAR imagery, could enhance the comprehensiveness of the IW

310   dataset.

The generated IW dataset potentially advances our understanding of IW characteristics in the northern SCS. By analyzing the spatial and temporal distributions of IWs based on the collected MODIS images, researchers can gain insights into the region's prevalent locations and seasonal variations of IW activity. This dataset also provides valuable information for studying the interactions between IWs and mesoscale ocean phenomena, such as eddies, facilitating further investigations into ocean dynamics (Li et al., 2016; Xie et al., 2016). The cyclone and anticyclone mesoscale eddies can cause vertical fluctuations in ocean temperature isopleths and accompany currents, affecting characteristics such as the amplitude and propagation direction of IWs. By analyzing the IW spatial and temporal information provided in this dataset, changes in IW characteristics after passing through different types of eddies can be statistically examined over longer timescales. Additionally, other dynamic phenomena, such as the intrusion of the Kuroshio current, also affect the generation and propagation of IWs in the SCS. Analyzing the statistical characteristics of IWs across different seasons and years can enhance understanding of how dynamic phenomena like the Kuroshio affect the IW generation and propagation process, ultimately improving the study and knowledge of multi-scale dynamic interactions in the SCS.

Moreover, the availability of this extensive IW dataset is crucial for advancing artificial intelligence oceanography studies. It serves as valuable ground truth data for validating IW generation or forecast models, allowing researchers to assess the performance of AI models by comparing their predictions with the dataset-provided IW locations. The dataset serves as a validation tool for numerical simulations (Gong et al., 2023), enabling researchers to refine and improve numerical models based on the observed IW distributions. This dataset can also serve as a benchmark for collaborative observations of IWs in the SCS with other satellite sensors or field measurements, thereby facilitating the construction of matched datasets to support IW research with artificial intelligence technologies.

Overall, the IW dataset presented in this paper significantly contributes to oceanography, providing researchers with a valuable resource for studying IW dynamics, validating AI models, and refining numerical simulations. This dataset is anticipated to stimulate further research and advancements in understanding the complex dynamics of oceanic IWs. Mooring observations can provide vertical structural information of IWs. By matching this dataset with mooring observation data and combining it with artificial intelligence technology, it is hoped to extend from two-dimensional sea surface information of IWs to a three-dimensional IW structure.

6. Data availability

The built IW dataset can be freely downloaded from http://dx.doi.org/10.12157/IOCAS.20240409.001 (Zhang and Li, 2024) or https://datapid.cn/CSTR:33685.11.IOCAS.20240409.001.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

Financial support

References

350

Alford, M. H., Peacock, T., MacKinnon, J. A., Nash, J. D., Buijsman, M. C., Centurioni, L. R., Chao, S. Y., Chang, M. H., Farmer, D. M., Fringer, O. B., Fu, K. H., Gallacher, P. C., Graber, H. C., Helfrich, K. R., Jachec, S. M., Jackson, C. R., Klymak, J. M., Ko, D. S., Jan, S., Johnston, T. M., Legg, S., Lee, I. H., Lien, R. C., Mercier, M. J., Moum, J. N., Musgrave, R., Park, J. H., Pickering, A. I., Pinkel, R., Rainville, L., Ramp, S. R., Rudnick, D. L., Sarkar, S., Scotti, A., Simmons, H. L., St Laurent,
355 L. C., Venayagamoorthy, S. K., Wang, Y. H., Wang, J., Yang, Y. J., Paluszkiewicz, T., and Tang, T. Y.: The formation and fate of internal waves in the South China Sea, Nature, 521, 65-69, 10.1038/nature14399, 2015.

Bai, X., Li, X., Lamb, K. G., and Hu, J.: Internal Solitary Wave Reflection Near Dongsha Atoll, the South China Sea, J. Geophys. Res.: Oceans, 122, 7978-7991, 10.1002/2017jc012880, 2017.

Bai, X., Liu, Z., Li, X., and Hu, J.: Generation sites of internal solitary waves in the southern Taiwan Strait revealed by MODIS
360 true-colour image observations, Int. J. Remote Sens., 35, 4086-4098, 10.1080/01431161.2014.916453, 2014.

Bao, S., Meng, J., Sun, L., and Liu, Y.: Detection of ocean internal waves based on Faster R-CNN in SAR images, J. Oceanol. Limnol., 38, 55-63, 10.1007/s00343-019-9028-6, 2019.

Cai, S., Xie, J., and He, J.: An Overview of Internal Solitary Waves in the South China Sea, Surv. Geophys., 33, 927-943, 10.1007/s10712-012-9176-0, 2012.

365 de Macedo, C. R., Koch-Larrouy, A., da Silva, J. C. B., Magalhães, J. M., Lentini, C. A. D., Tran, T. K., Rosa, M. C. B., and Vantrepotte, V.: Spatial and temporal variability in mode-1 and mode-2 internal solitary waves from MODIS-Terra sun glint off the Amazon shelf, Ocean Sci., 19, 1357-1374, 10.5194/os-19-1357-2023, 2023.

Dong, D., Yang, X. F., Li, X. F., and Li, Z. W.: SAR Observation of Eddy-Induced Mode-2 Internal Solitary Waves in the South China Sea, IEEE Trans. Geosci. Remote Sensing, 54, 6674-6686, 10.1109/Tgrs.2016.2587752, 2016.

370 Furtney, S., Romeiser, R., and Graber, H. C.: Automated retrieval of internal wave phase speed and direction from pairs of SAR images with different look directions, Remote Sens. Environ., 305, 114084, 2024.

Gong, Y., Chen, X., Xu, J., Xie, J., Chen, Z., He, Y., and Cai, S.: An internal solitary wave forecasting model in the northern South China Sea (ISWFM-NSCS), Geosci. Model Dev., 16, 2851-2871, 10.5194/gmd-16-2851-2023, 2023.

Guo, C. and Chen, X.: A review of internal solitary wave dynamics in the northern South China Sea, Prog. Oceanogr., 121, 7-
375 23, 10.1016/j.pocean.2013.04.002, 2014.

Haury, L. R., Briscoe, M. G., and Orr, M. H.: Tidally generated internal wave packets in Massachusetts Bay, Nature, 278, 312-317, 10.1038/278312a0, 1979.

Hu, B. L., Meng, J. M., Sun, L. N., and Zhang, H.: A Study on Brightness Reversal of Internal Waves in the Celebes Sea Using Himawari-8 Images, Remote Sens., 13, 10.3390/Rs13193831, 2021.

380 Jia, T., Liang, J. J., Li, X. M., and Sha, J.: SAR Observation and Numerical Simulation of Internal Solitary Wave Refraction and Reconnection Behind the Dongsha Atoll, J. Geophys. Res.: Oceans, 123, 74-89, 10.1002/2017jc013389, 2018.

Jia, Y., Tian, Z., Shi, X., Liu, J. P., Chen, J., Liu, X., Ye, R., Ren, Z., and Tian, J.: Deep-sea sediment resuspension by internal solitary waves in the northern South China Sea, Sci. Rep., 9, 12137, 10.1038/s41598-019-47886-y, 2019.

Kurekin, A. A., Land, P. E., and Miller, P. I.: Internal Waves at the UK Continental Shelf: Automatic Mapping Using the
385 ENVISAT ASAR Sensor, Remote Sens., 12, 2476, 10.3390/rs12152476, 2020.

Li, Q., Wang, B., Chen, X., Chen, X., and Park, J. H.: Variability of nonlinear internal waves in the South China Sea affected by the Kuroshio and mesoscale eddies, J. Geophys. Res.: Oceans, 121, 2098-2118, 10.1002/2015jc011134, 2016.

Li, X., Jackson, C. R., and Pichel, W. G.: Internal solitary wave refraction at Dongsha Atoll, South China Sea, Geophys. Res. Lett., 40, 3128-3132, 10.1002/grl.50614, 2013.

390 Li, X., Zhao, Z., and Pichel, W. G.: Internal solitary waves in the northwestern South China Sea inferred from satellite images, Geophys. Res. Lett., 35, 10.1029/2008gl034272, 2008.

Li, X., Zhou, Y., and Wang, F.: Advanced Information Mining from Ocean Remote Sensing Imagery with Deep Learning, Journal of Remote Sensing, 2022, 1-4, 10.34133/2022/9849645, 2022.

Li, X., Liu, B., Zheng, G., Ren, Y. B., Zhang, S. S., Liu, Y. J., Gao, L., Liu, Y. H., Zhang, B., and Wang, F.: Deep-learning-
395 based information mining from ocean remote-sensing imagery, Natl. Sci. Rev., 7, 1584-1605, 10.1093/nsr/nwaa047, 2020.

Liang, J., Li, X.-M., Sha, J., Jia, T., and Ren, Y.: The Lifecycle of Nonlinear Internal Waves in the Northwestern South China Sea, J. Phys. Oceanogr., 10.1175/jpo-d-18-0231.1, 2019.

Liu, A. K. and Hsu, M. K.: Internal wave study in the South China Sea using Synthetic Aperture Radar (SAR), Int. J. Remote Sens., 25, 1261-1264, 10.1080/01431160310001592148, 2004.

400 Liu, B., Li, X., and Zheng, G.: Coastal inundation mapping from bitemporal and dual-polarization SAR imagery based on deep convolutional neural networks, J. Geophys. Res.: Oceans, 124, 9101-9113, 10.1029/2019jc015577, 2019.

Liu, B., Yang, H., Zhao, Z., and Li, X.: Internal solitary wave propagation observed by tandem satellites, Geophys. Res. Lett., 41, 2077-2085, 10.1002/2014GL059281, 2014.

Liu, T. Y. and Abernathey, R.: A global Lagrangian eddy dataset based on satellite altimetry, Earth System Science Data, 15,
405 1765-1778, 10.5194/essd-15-1765-2023, 2023.

Liu, T. Y., He, Y. H., Zhai, X. M., and Liu, X. H.: Diagnostics of Coherent Eddy Transport in the South China Sea Based on Satellite Observations, Remote Sens., 14, 10.3390/rs14071690, 2022.

Liu, T. Y., Xu, J. X., He, Y. H., Lü, H. B., Yao, Y., and Cai, S. Q.: Numerical simulation of the Kuroshio intrusion into the South China Sea by a passive tracer, Acta Oceanologica Sinica, 35, 1-12, 10.1007/s13131-016-0930-x, 2016.

410 Ma, Y. T., Meng, J. M., Sun, L. N., and Ren, P.: Oceanic Internal Wave Signature Extraction in the Sulu Sea by a Pixel Attention U-Net: PAU-Net, IEEE Geosci. Remote Sens. Lett., 20, 10.1109/Lgrs.2022.3230086, 2023.

Magalhaes, J. M., da Silva, J. C. B., and Buijsman, M. C.: Long lived second mode internal solitary waves in the Andaman Sea, Sci. Rep., 10, 10234, 10.1038/s41598-020-66335-9, 2020.

Magalhaes, J. M., da Silva, J. C. B., Nolasco, R., Dubert, J., and Oliveira, P. B.: Short timescale variability in large-amplitude
415 internal waves on the western Portuguese shelf, Cont. Shelf Res., 246, 10.1016/j.csr.2022.104812, 2022.

Pan, J., Jay, D. A., and Orton, P. M.: Analyses of internal solitary waves generated at the Columbia River plume front using SAR imagery, J. Geophys. Res.: Oceans, 112, 10.1029/2006jc003688, 2007.

Ramp, S. R., Yang, Y. J., Chiu, C.-S., Reeder, D. B., and Bahr, F. L.: Observations of shoaling internal wave transformation over a gentle slope in the South China Sea, Nonlinear Proc. Geoph., 29, 279-299, 10.5194/npg-29-279-2022, 2022a.

420 Ramp, S. R., Yang, Y. J., Jan, S., Chang, M. H., Davis, K. A., Sinnett, G., Bahr, F. L., Reeder, D. B., Ko, D. S., and Pawlak, G.: Solitary waves impinging on an Isolated tropical reef: arrival patterns and wave transformation under shoaling, J. Geophys. Res.: Oceans, 127, 10.1029/2021jc017781, 2022b.

Sun, L., Zhang, J., and Meng, J.: Study on the propagation velocity of internal solitary waves in the Andaman Sea using Terra/Aqua-MODIS remote sensing images, J. Oceanol. Limnol., 39, 2195-2208, 10.1007/s00343-020-0280-6, 2021.

425 Tao, M., Xu, C., Guo, L., Wang, X., and Xu, Y.: An Internal Waves Data Set From Sentinel-1 Synthetic Aperture Radar Imagery and Preliminary Detection, Earth Space Sci., 9, e2022EA002528, 2022.

Wang, H. and Li, X.: DeepBlue: Advanced convolutional neural network applications for ocean remote sensing, IEEE Geosc. Rem. Sen. M., 10.1109/Mgrs.2023.3343623, 2023.

430 Xie, J., He, Y., Lü, H., Chen, Z., Xu, J., and Cai, S.: Distortion and broadening of internal solitary wavefront in the northeastern South China Sea deep basin, Geophys. Res. Lett., 43, 7617-7624, 10.1002/2016gl070093, 2016.

Xu, J., He, Y., Chen, Z., Zhan, H., Wu, Y., Xie, J., Shang, X., Ning, D., Fang, W., and Cai, S.: Observations of different effects of an anti-cyclonic eddy on internal solitary waves in the South China Sea, Prog. Oceanogr., 188, 102422, 10.1016/j.pocean.2020.102422, 2020.

Zhang, M., Wang, J., Chen, X., Mei, Y., and Zhang, X.: An experimental study on the characteristic pattern of internal solitary 435 waves in optical remote-sensing images, Int. J. Remote Sens., 40, 7017-7032, 10.1080/01431161.2019.1597308, 2019.

Zhang, S., Li, X., and Zhang, X.: Internal Wave Signature Extraction from SAR and Optical Satellite Imagery Based on Deep Learning, IEEE Trans. Geosci. Remote Sensing, 61, 10.1109/TGRS.2023.3258189, 2023.

Zhang, X., Wang, H., Wang, S., Liu, Y., Yu, W., Wang, J., Xu, Q., and Li, X.: Oceanic internal wave amplitude retrieval from satellite images based on a data-driven transfer learning model, Remote Sens. Environ., 272, 112940, 440 10.1016/j.rse.2022.112940, 2022.

Zhao, W., Huang, X., and Tian, J.: A new method to estimate phase speed and vertical velocity of internal solitary waves in the South China Sea, Journal of Oceanography, 68, 761-769, 10.1007/s10872-012-0132-x, 2012.

Zhao, Z., Liu, B., and Li, X.: Internal solitary waves in the China seas observed using satellite remote-sensing techniques: a review and perspectives, Int. J. Remote Sens., 35, 3926-3946, 10.1080/01431161.2014.916442, 2014.

445 Zhao, Z., Klemas, V., Zheng, Q., Li, X., and Yan, X.: Estimating parameters of a two-layer stratified ocean from polarity conversion of internal solitary waves observed in satellite SAR images, Remote Sens. Environ., 92, 276-287, 10.1016/j.rse.2004.05.014, 2004.

Zheng, Q., Yuan, Y., Klemas, V., and Yan, X.-H.: Theoretical expression for an ocean internal soliton synthetic aperture radar image and determination of the soliton characteristic half width, J. Geophys. Res.: Oceans, 106, 31415-31423, 450 10.1029/2000jc000726, 2001.

Zheng, Y. G., Zhang, H. S., and Wang, Y. Q.: Stripe detection and recognition of oceanic internal waves from synthetic aperture radar based on support vector machine and feature fusion, Int. J. Remote Sens., 42, 6710-6728, 10.1080/01431161.2021.1943040, 2021.