JGR Oceans

RESEARCH ARTICLE
10.1029/2023JC020686

A Novel Sea State Classification Scheme of the Global CFOSAT Wind and Wave Observations

Huimin Li[1,2], Bertrand Chapron[3], Douglas Vandemark[4], Wenming Lin[1,2], Danièle Hauser[5], Yijun He[1,2], and Fabrice Collard[6]

[1]School of Marine Sciences, Nanjing University of Information Science & Technology, Nanjing, China, [2]Key Laboratory of Space Ocean Remote Sensing and Application, Ministry of Natural Resources, Beijing, China, [3]Ifremer, University Brest, CNRS, IRD, Laboratoire d'Océanographie Physique et Spatiale (LOPS), IUEM, Brest, France, [4]Ocean Processes Analysis Laboratory, University of New Hampshire, Durham, NH, USA, [5]Laboratoire Atmosphère, Milieux, Observations Spatiales (LATMOS), Université Paris-Saclay, UVSQ, Centre National de la Recherche Scientifique (CNRS), Sorbonne Université, Guyancourt, France, [6]OceanDataLab, Locmaria-Plouzane, France

Abstract Ocean waves are essential elements across the air-sea interface, regulating momentum and energy transfer. The mixture of wind sea and ocean swell coupled with surface winds results in diverse sea state conditions that modify the local air-sea interaction. Previous classifications of wind waves and swells are mostly binary that are insufficient to represent the complexity of sea states. In this study, we utilize wind and wave measurements from the China-France Oceanography Satellite (CFOSAT) to construct an observational wind-wave ensemble. Four key parameters: wind speed, significant wave height, inverse wave age, and spectral width are selected out of six variables based on their correlations. Employing the unsupervised learning of k-means clustering, global sea states are categorized into six distinct classes. These classes, characterized by unique centroids and separated in the feature space, represent specific wind regimes and degrees of wave development. Global occurrence highlights that each sea state is region-specific, bridging the spatial gap of swell and wind sea dominated areas, respectively. This new grouping scheme complements the traditional wind sea and/or swell classification by resolving the diversity of wave regimes. The six-class classification enables us to identify transitional states and hybrid conditions that may have been overlooked in the binary classification scheme, which shall help investigate the impact of ocean waves on the air-sea interaction under varying sea states.

Plain Language Summary Ocean waves are vital in the interaction between the ocean and the atmosphere, affecting the exchange of energy and momentum. The China-France Oceanography Satellite (CFOSAT) can provide concurrent observations of wind vectors and ocean wave directional spectrum. This study selects four key wind and wave parameters from CFOSAT measurements: wind speed, significant wave height, inverse wave age, and spectral width. A type of unsupervised learning technique, k-means clustering, is applied to group these parameters and classify different sea states. Through this method, six distinct classes of sea states are obtained each representing a specific combination of wind and wave characteristics. These sea state classes are unique and independent when looked at together in a four-dimensional feature space. However, when considering just one variable, they can overlap. This approach shows that each sea state has particular regional variations, helping bridge the gap between regions dominated by swell and those dominated by wind-generated waves.

Key Points:

- The concurrent wind and wave variables derived from China-France Oceanography Satellite (CFOSAT) observations are analyzed to create a 4-dimensional wind-wave ensemble
- Six sea state groups are classified by k-means clustering to characterize different regimes of wind and wave conditions
- Global frequency of occurrence of these six sea state classes provides greater insight into the regional dominance of wind sea and swell

Correspondence to:
B. Chapron,
bertrand.chapron@ifremer.fr

Citation:
Li, H., Chapron, B., Vandemark, D., Lin, W., Hauser, D., He, Y., & Collard, F. (2024). A novel sea state classification scheme of the global CFOSAT wind and wave observations. *Journal of Geophysical Research: Oceans*, *129*, e2023JC020686. https://doi.org/10.1029/2023JC020686

Received 11 DEC 2023
Accepted 7 OCT 2024

Author Contributions:
Conceptualization: Bertrand Chapron
Data curation: Wenming Lin
Formal analysis: Huimin Li
Funding acquisition: Yijun He
Methodology: Bertrand Chapron, Douglas Vandemark, Wenming Lin
Supervision: Douglas Vandemark, Danièle Hauser, Yijun He, Fabrice Collard
Validation: Huimin Li
Writing – original draft: Huimin Li
Writing – review & editing: Bertrand Chapron, Douglas Vandemark, Wenming Lin, Danièle Hauser, Yijun He, Fabrice Collard

1. Introduction

Ocean surface waves generated by the surface wind forcing form the interface between the atmosphere and the underlying ocean. They are well acknowledged for their influence on the momentum, heat and mass flux transfer, being essential components in global atmosphere-ocean interactions and climate models (Cavaleri et al., 2012). Information on ocean waves is also important for the design of offshore infrastructures and the safety of sea-going activities, among others (Janssen et al., 2003). Many aspects of ocean wave related dynamics have been studied, from the mechanism of wave generation (Donelan et al., 2006; Miles, 1957), to their impact on the air-sea fluxes (Donelan et al., 1997; Shimura et al., 2020), to the global sea state climate (Young et al., 2011; X.-M. Li, 2016; Lobeto et al., 2022).

Ocean waves are usually categorized into two types in terms of their coupling to local winds: wind seas and swells. Wind sea is locally generated, under growth or in equilibrium with the local wind (Kinsman, 1965). Ocean

swell are wave systems originating from distant storms that outrun their generation area and radiate across the ocean basins. Over the long-distance propagation of thousands of kilometers, swell systems are almost not coupled to the local winds and barely extract momentum (Ardhuin et al., 2009; Collard et al., 2009; Snodgrass et al., 1966). Sea state is often a mixture of these two types of ocean waves that may exhibit the superposition of multiple swell systems with a local wind sea. Growing wind waves might extract momentum from the atmosphere, while swell waves contribute to modifying the momentum exchange in ways depending on their relative speed versus near surface wind speed (Donelan et al., 1997). It has been reported that upward momentum flux from the ocean to the atmosphere can occur under the conditions when swell propagates faster than the local wind speed (Ardhuin et al., 2010; Grachev & Fairall, 2001). To develop a climatology of wind–wave interactions that account for both wind waves and swell systems, it is therefore of great interest to monitor but also categorize the sea states.

The common approach for sea state classification is to separate wind sea dominated cases from swell dominated cases using a method based either on the simple relationship between significant wave height and wind speed that marks the limit of fully developed wind waves (G. Chen et al., 2002; Jiang & Chen, 2013) or the inverse wave age parameter (ratio of wind speed to dominant wave phase velocity) (Hanley et al., 2010). Chen et al. (2002) pioneered efforts in this regard using the collocated scatterometer-derived U_{10} and altimeter-estimated significant wave height (SWH and H_s used interchangeably hereinafter) (G. Chen et al., 2002). The global pattern of wind sea and swell dominance was portrayed, building on a prescribed wind–wave relation for fully developed seas, that is, H_s less than the fully developed limit is attributed to a growing wind sea while H_s above this limit is attributed to a swell system. An index to quantify the degree of swell dominance is defined based on the swell energy and the wind sea energy to characterize the spatial extent of swell-dominated regions. Three tongue-shaped "swell pools" were identified in the tropics over the regions with a probability of swell dominance higher than 95%. They persist at low latitudes throughout the year, in contrast to the later defined "seasonal swell pools" that are only present in hemispheric summer based on 10-year collocation data set between scatterometer and altimeter measurements (Jiang & Chen, 2013). Recently, a probability-based algorithm was proposed to identify the swell wave height from the JASON satellite in terms of the local wind-wave relation predefined based on the spectral energy partition of buoy directional wave spectrum (Gao et al., 2023). Chen et al. (2020) also computed the swell index based on the wave energy from the directional wave spectrum and obtained similar patterns with that based on numerical model results (S. Chen et al., 2020). Lobeto et al. (2022) used clustering techniques applied to directional wave spectra from numerical models to analyze the present and projected climatologies based on a seven-member wave climate projection ensemble. Such a classification helps provide a comprehensive understanding of the projected sea state changes within the context of global warming.

The classical separation scheme based on the (U_{10}, H_s) relation usually takes the fully-developed ocean waves as a reference to classify the wind-wave and swell dominance. In reality, the local wind sea, that is, associated with the local wind conditions, corresponds to different stages of wave development. This wave maturity is typically characterized by the inverse wave age (IWA). Various IWA forms have appeared in the literature, including U_{10}/c_p, $U_{10}\cos\phi/c_p$, U_*/c_p and $U_*\cos\phi/c_p$, where U_* denotes the surface friction velocity, c_p the phase velocity of peak waves and ϕ is the relative direction between surface wind and peak wave propagation (Högström et al., 2011). Whichever is used, IWA derivation requires the peak waves to be determined and the phase velocity evaluated. A common practical criterion to distinguish wind sea from swell considers the phase velocity of a wave scale being 1.2 times the local wind speed (Elfouhaily et al., 1997); in other words, IWA equal to 0.84. Given difficulties obtaining the directional wave spectra at an extended spatial scale, the IWA-involved wave classification studies are mostly carried out using the numerical model reanalysis data set. Spatial patterns of swell probability in terms of both H_s and wind sea energy (Semedo et al., 2011) were reported consistent with the satellite observations (G. Chen et al., 2002). Such frequent swell presence across the globe is also evidenced by the results that IWA is mostly smaller than 0.6 in the open ocean, lower than the value of 0.84 for the fully developed sea state (Hanley et al., 2010).

These classification methods mostly lead to a binary categorization of wind sea dominance and swell dominance although it has been broadly acknowledged that the simultaneous presence of both wind sea and swell is very common (G. Chen et al., 2002; Semedo et al., 2011; Lobeto et al., 2022). The H_s versus U_{10} relationship is not unique, as local conditions might vary by an order of magnitude for a given wind speed. The previously reported classification schemes of wind sea and swell based on prescribed hard-limit criteria of H_s versus U_{10} or inverse

wave age, while useful, are insufficient to represent the complexity of sea states. Analyses into multiple classes are necessary to capture the diverse range of oceanic conditions more accurately. Lobeto et al. (2022) proposed a method to analyze the climatology of the full two-dimensional wave spectra that requires first wave partitioning and reconstruction of the wave spectra from estimated partition parameters. Here, we propose an alternative method, based on the analysis of 4 parameters which characterize the wave spectra and wind speed. It is thus intermediate between the simple classification with two variables (H_s vs. U_{10} or IWA) and the more complex method in Lobeto et al. (2022). This classification provides a method of distinguishing between different levels of wave development, from calm seas to higher waves, and to identify transitional states that lie between conventional categories.

We take advantage of the newly concurrent wind and wave measurements acquired by the China-France Oceanography Satellite (CFOSAT) (Hauser et al., 2021), to classify the sea states on a data-driven basis. CFOSAT can provide the wind vector (speed and direction) and directional wave spectrum, from which wave-related descriptors could be readily derived. Though the directional wave spectrum could provide a full description of sea state, the integrated wave parameters have been proven useful representing the mean conditions (Lobeto et al., 2022). A wind and wave ensemble composed of wind speed, H_s, IWA and the spectral width estimated from the CFOSAT product is created. We then use the widely applied k-means clustering algorithm to identify the wind-wave classes, each with a certain degree of correlation of wind and waves characteristics but distinct between classes. A 3-year data set acquired by CFOSAT from May 2019 to December 2022 is included here to quantify the global occurrence of frequency of each class. For the organization of this study, data and methods are presented in Section 2. Results are given in Section 3, followed by Discussion in Section 4. Summary and perspectives are finally presented in Section 5.

2. CFOSAT Wind and Wave Ensemble

2.1. Data Description

Despite the broad use of spaceborne instruments for ocean monitoring, CFOSAT is the first of its kind carrying two sensors to provide concurrent wind vectors and directional wave spectrum observations on a global scale with a revisit cycle of 13 days. The rotating fan-beam scatterometer (SCAT) can measure the surface wind vectors at a spatial resolution of 25 km (Liu et al., 2020). Note that the SCAT wind product experienced a processor change effective on 27 May 2020 from version 2.0 to version 3.0. Here, we use the neutral equivalent winds at 10 m height (U_{10}) in the level-2 product from May 2019 to December 2022 with data from both versions included. The surface wave investigation and monitoring (SWIM) instrument is a wave spectrometer operating at nadir (an advanced altimeter) and at near-nadir incidence angles, that is, $(0°, 10°)$ with a step of 2°. Such a real aperture radar with rotating near-nadir beams has been demonstrated to be capable of observing the directional wave spectrum (Hauser et al., 2001). Each of the rotating beams have an azimuthal resolution of 7.5° with an original slant range sampling of 0.47 m. After the successful launch of CFOSAT in October 2018, the SWIM instrument processor facility (IPF) experienced a significant update (version 5.1.2) in September 2020. All data acquired earlier reprocessed using this version are included here, along with the products processed in near-real time with the same 5.1.2 version from September 2020 to December 2022. This updated IPF takes an improved speckle noise correction method based on an empirical model derived directly from the observations and an upgraded modulation transfer function to match the nadir measured SWH (Hauser et al., 2021). It should be noted that SWIM products of this version are not yet quality controlled in terms of rain contamination. Also, the presence of parasitic peaks at low wavenumbers due to either scene inhomogeneity or spectral energy amplification in obtaining the wave height spectrum might influence the wave parameters (Xu et al., 2022). In this study, wave variables of SWH, peak wavelength and peak wave direction derived from the 10° spectral beam are directly taken from the level-2 products given its best quality of wave measurements (Hauser et al., 2021). The peak wavelength and direction provided in the SWIM products are estimated from the wave slope spectrum, and thus are only marginally affected by the parasitic peak problem which appears mainly on the wave height spectra. Each box measurement is available at a spatial resolution of 70 km by 90 km. Note that these wave parameters are calculated based on the whole wave spectrum. To possibly mitigate the impact of parasitic peaks, the 1% highest peak wavelengths are excluded from the subsequent analysis.

Collocations of the SCAT winds and the SWIM wave parameters are carried out as follows. Given their unmatched spatial resolution, all the wind cells within 50 km from the center of a given wave cell are averaged. Such

a wind vector is used to compose the wind-wave ensemble. Note that the time difference between wind and wave observations is generally within a few minutes, sufficiently small to be neglected for the present purpose. This ends up with 9,007,641 collocation points with a mean wind speed of 7.5 $m \cdot s^{-1}$ and a mean SWH of 2.7 m, respectively. Across the open ocean, the data are evenly distributed with more than 1,100 points in most of the grids at 2 ° × 2° size (not shown). The uniform distribution of data density ensures that the global data are equally considered in the clustering analysis, avoiding significant biases toward specific regions. Note that the data points with a wind speed less than 3 $m \cdot s^{-1}$ account for only 4.8% of the total data set. On the other hand, there are only 0.001% of points with a SWH below 0.20 m, which is considered the SWH of a fully developed wind sea using the Pierson-Moskowitz (PM) spectrum of $0.22U_{10}^2/g$ for a 3 $m \cdot s^{-1}$ U_{10} (where g is the gravitational acceleration). This roughly indicates the common presence of ocean swell.

In addition to SWH, the following wave parameters are also calculated. The inverse wave age is calculated via $IWA = |U_{10} \cdot \cos(\phi_p - \phi_U)/c_p|$, where ϕ_U denotes the wind direction. ϕ_p and c_p are the wave direction and phase velocity of the peak waves, respectively. Note the absolute value of IWA is taken here for simplicity by neglecting the 180° ambiguity in the wave propagation direction from SWIM. This means that the conditions where swell propagating at an angle larger than ±90° from the wind direction cannot be resolved. In addition, using the wave parameters provided in the SWIM products we estimate the following additional parameters: mean surface velocity (MSV), mean square surface slope (MSS) and spectral width. MSV is defined as the variance of wave orbital velocities, obtained by $MSV = \int_0^{2\pi}\int_{2\pi/500}^{2\pi/22}\omega^2 \cdot F(k, \phi)\, k\, dk\, d\phi$. This definition of MSV is similar to that by Stopa et al. (2015) and has been shown to be associated with the azimuth cutoff wavelength in SAR wave measurements. It is related to the second-order moment of the frequency spectrum, complementary to the zeroth-order moment used to estimate SWH. The total MSS is computed by $MSS = \int_0^{2\pi}\int_{2\pi/500}^{2\pi/22}k^2 \cdot F(k, \phi)\, k\, dk\, d\phi$. The spectral width is defined as $\varepsilon_v = \sqrt{\frac{m_0 m_2}{m_1^1} - 1}$ where m_n is the nth moment of the directional wave number spectrum (Longuet-Higgins, 1975). This spectral width can be viewed as a measure of the peak separation between wind sea and swell for a given wave system. Based on a simulation using the JONSWAP spectrum, we found that in the case of unimodal spectra, the spectral width appears primarily sensitive to peak wavelength rather than to the specific type of spectrum chosen (narrow swell-type or broader wind sea). However, for bimodal spectra, combining swell and wind sea components, the spectral width significantly decreases compared to swell-only spectra, especially when the swell spectrum has the largest energy peak. Note that it differs from the definition used by Saulnier et al. (2011) and by Le Merle et al. (2021) in that the peakedness parameter shows much weaker variation relative to either wind speed or inverse wave age compared to the above definition here.

CFOSAT wind and wave products have been demonstrated to be of high quality and suitable for scientific exploitation (Aouf et al., 2021; Hauser et al., 2021). To ensure that the reprocessed data also meet the accuracy requirements, an assessment of U_{10} and SWH is carried out using the multi-source, multi-platform observations from the Copernicus Marine Service (CMEMS) that undergo consistent quality control. The CFOSAT-measured U_{10} and SWH are collocated with CMEMS data within 70 km in space and 30 min in time, resulting in 48,080 and 8,413 points, respectively. Note that all buoy observations within 20 km of the coastline are excluded from the collocations. Figures 1a and 1b show that the wind measuring buoys have a better coverage than the wave buoys, causing a disparity in the number of collocated wind and wave points. Since most of the buoys measure wind speed around 4 m, a logarithmic wind profile formula is applied to convert to neutral wind speeds at 10 m assuming neutral conditions. The formula used is $U_{10} = U_z \times \left[\frac{\ln(10/z_0)}{\ln(z/z_0)}\right]$, where U_{10} is the wind speed at 10 m, U_z is the measured wind speed at height z (usually 4 m), and z_0 is the roughness length, typically taken as 0.0002 m over the open sea (Paulson, 1970). The comparison between the scatterometer-measured U_{10} and the *in situ* wind speed is plotted in Figure 1c, showing a mean bias of −0.40 $m \cdot s^{-1}$ and root-mean-square-error (RMSE) of 1.21 $m \cdot s^{-1}$.The best agreement is observed for the wind speed range of 5–15 $m \cdot s^{-1}$ with U_{10} overestimated at low winds. This might be due to the reduced performance of the wind direction ambiguity removal algorithm at low sea state (Ebuchi et al., 2002; Yang & Zhang, 2019) and also probably due to the low signal-to-noise ratio at low wind (Liu et al., 2020). SWIM SWH relative to the buoy measurements is given in Figure 1d with a mean bias of 0.03 m and RMSE of 0.31 m. Most of the data pairs scatter around the one-to-one line and do not exhibit a skewed trend like wind speed. The performance of U_{10} and SWH is consistent with previous validation efforts (Hauser

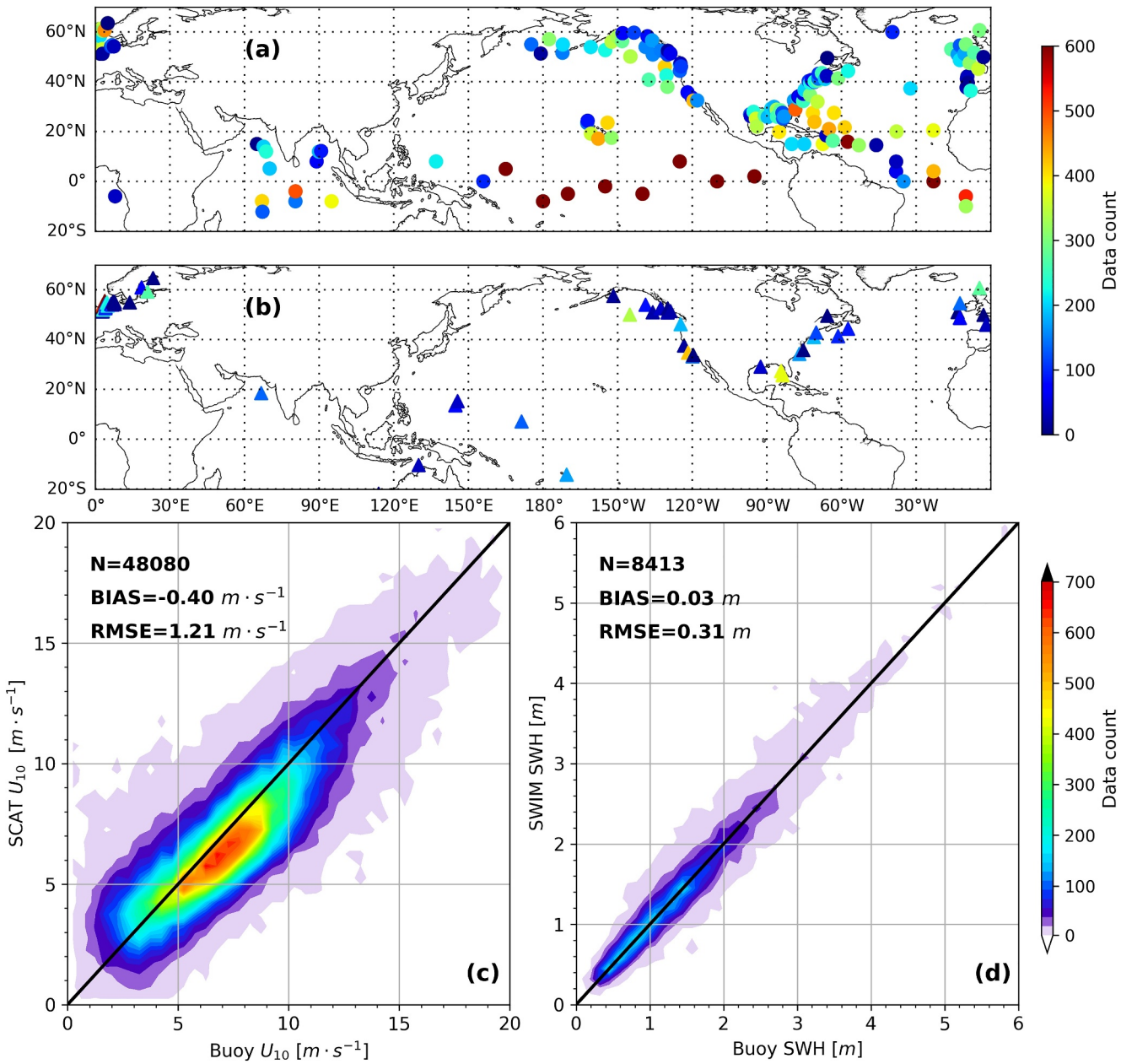

Figure 1. Spatial location of collocated buoys for (a) wind and (b) wave measurements validation. Color denotes the data count at each buoy location. Q-Q plot of (c) scatterometer observed wind speed relative to the buoy measurements; (d) SWIM derived significant wave height relative to the buoy. The color scale is the data count at a wind speed bin of 0.5 $m \cdot s^{-1}$ and SWH bin of 0.125 m, respectively. CFOSAT products acquired from May 2019 to December 2022 are included.

et al., 2021; Liu et al., 2020). Note that the collocated data points pole-ward of 60° latitude are excluded to minimize the potential impact of sea ice on wave variables.

2.2. Correlation Analysis

Prior to developing a clustering algorithm, it is necessary to study the correlation between the 6 parameters defined above (U_{10}, SWH, IWA, MSV, MSS, ε_v) in order to choose a subset of parameters which are less correlated. The correlation coefficients of sea state variables are computed and presented in Figure 2a. U_{10} has a positive correlation with SWH, IWA, MSV and MSS, and a weakly negative correlation with the spectral width (ε_v). On one hand, SWH has a high correlation with both MSS and MSV, as evidenced by Figures 2a and 2b. Given a SWH value, MSV varies within a narrow range of values. This also holds for MSS (not shown here). On

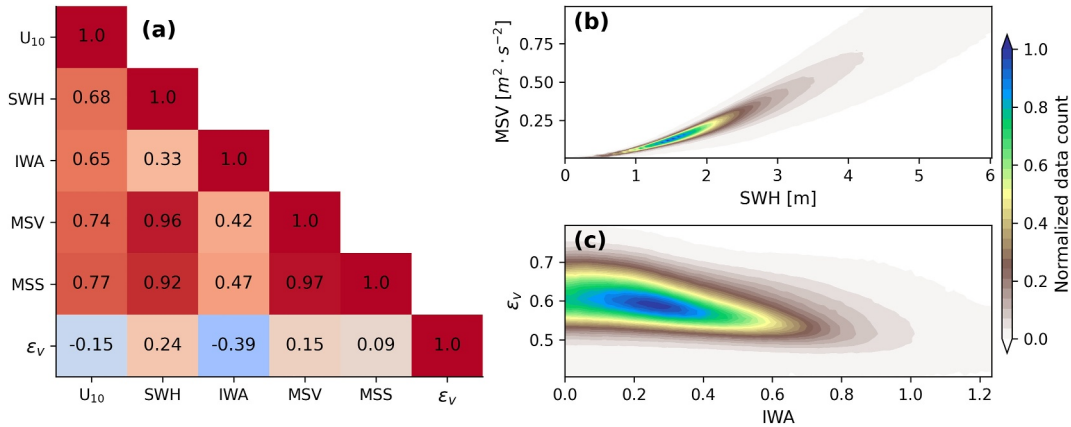

Figure 2. (a) Correlation matrix between the six sea state parameters. Density plot of (b) MSV versus SWH and (c) spectral width ε_v versus IWA to illustrate their correlation coefficient.

the other hand, SWH is weakly related to IWA and to the spectral width with a correlation coefficient of 0.33 and 0.24, respectively. IWA has a negative correlation with the spectral width, implying that their variation trends are opposite, as shown in Figure 2c. Smaller IWA indicates the dominance of swell systems that are generated by remote storms, usually along with low winds. Under such conditions, the wave peaks of swell and wind sea are well separated in the spectral domain, resulting in a higher spectral width. In contrast, a larger IWA is associated with wind sea dominance and closer wave peaks of wind sea and swell, corresponding to a smaller spectral width. This indicates that the spectral width is mainly influenced by IWA, while the scatter observed at given IWA may be attributed to the variety of wave spectra, whether unimodal or bimodal. Based on this correlation analysis, the U_{10}, SWH, IWA and ε_v are selected as the four parameters to create a wind and wave ensemble data set. MSS and MSV are not retained in the following analysis because of their strong correlation with SWH.

2.3. Mean and Variability of Sea State Variables

The compiled global averages of wind and wave observations are shown in Figure 3. The wind speed distribution in Figure 3a conforms with previous satellite measurements (Desbiolles et al., 2017; Risien & Chelton, 2008). The

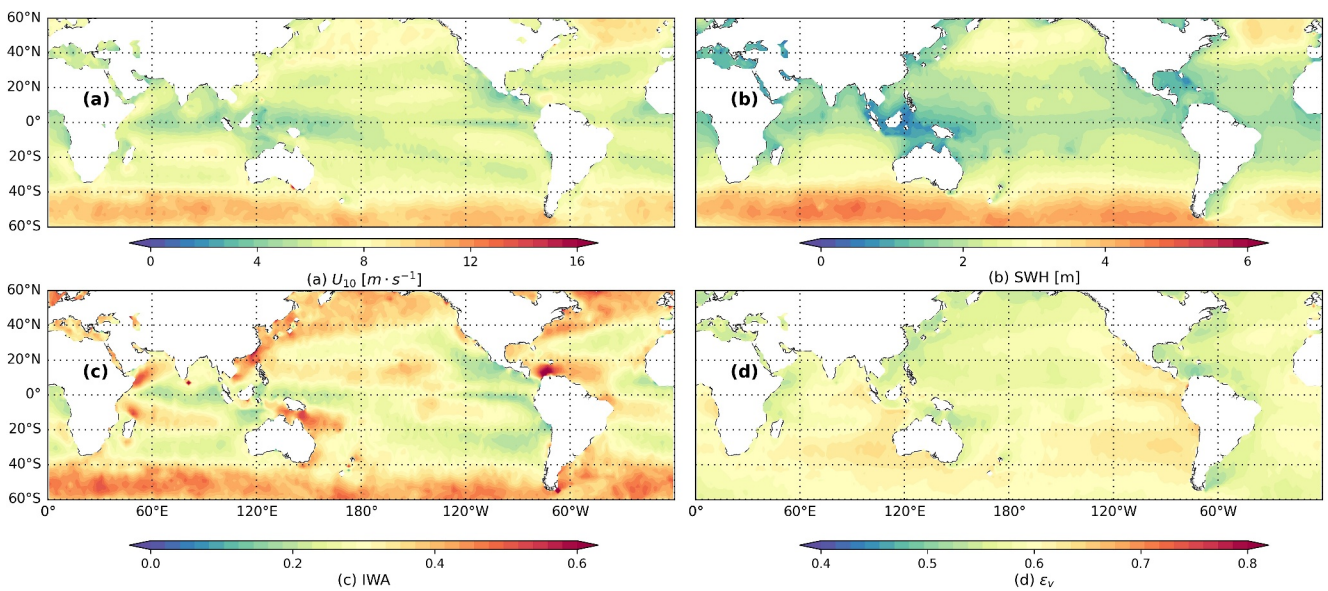

Figure 3. Global average of the wind and wave parameter for (a) U_{10}; (b) SWH; (c) IWA and (d) ε_v. The maps are compiled at the latitude/longitude bin of 2° from May 2019 to December 2022. Note that the color bar for each metric is annotated in the figure bottom.

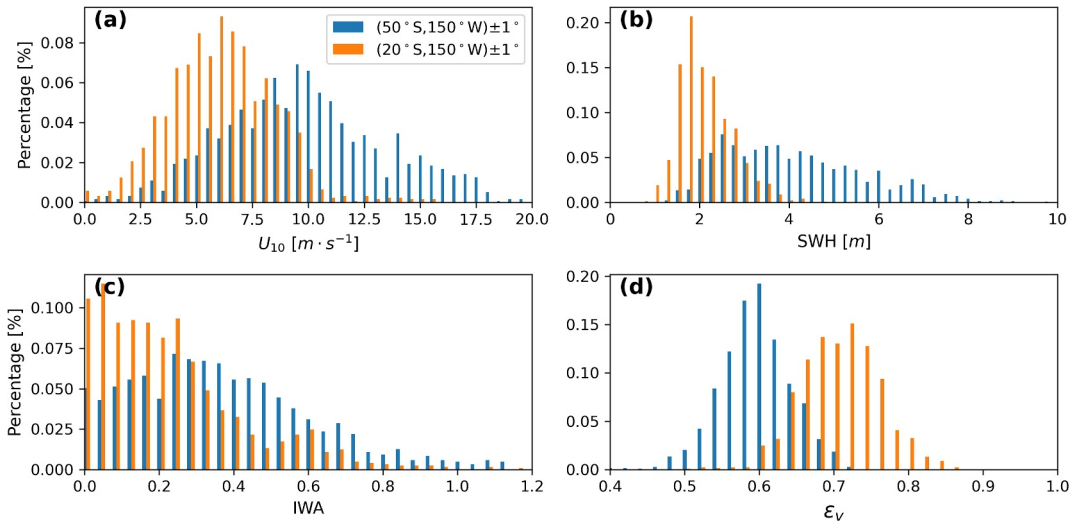

Figure 4. Histogram of (a) U_{10}; (b) SWH; (c) IWA and (d) ε_v at two chosen locations of (20°S,150°W) ± 1° and (50°S,150° W) ± 1°. The total number of data points is 1,204 and 1,191 for these two spatial boxes, respectively.

meridional variation of U_{10} is clear with low to medium winds in the tropics and high winds across the storm tracks in both hemispheres. The averaged U_{10} difference between the two extratropics (latitude of 40°-60°) is around 1 $m \cdot s^{-1}$ in the Pacific Ocean, and larger than the 0.1 $m \cdot s^{-1}$ in the Atlantic Ocean. The meridional pattern of SWH in Figure 3b is similar to that of wind speed with a clear variation across different ocean regions. Ocean waves generated in the Southern Ocean propagate northward and create high SWH values in the Indian and Pacific Oceans. In the Atlantic Ocean sector, SWH is lower across the Southern Hemisphere storm tracks, possibly due to the limited fetch influence by the Drake passage. This could be expected from the relationship of SWH versus fetch distance. Taking the relation of $H_s = 4.0\sqrt{1.67 \times 10^{-7} \times U_{10}^2/g \times F}$ with g being the gravitational acceleration and F the fetch length (Hasselmann et al., 1973), a typical 10 $m \cdot s^{-1}$ shall produce a 1.6 m SWH for a fetch length of 100 km and 3.7 m for $F = 500$km.

Figure 3c shows that the relatively young sea states characterized by the largest IWA are mainly found in areas of high wind conditions along the storm tracks. Ocean waves generated by distant storms and propagating out of their original locations become swell with no/weak coupling to the local winds (Collard et al., 2009; Hell et al., 2021; Kudryavtsev et al., 2021). When local wind remains low, the IWA remains small even in the presence of wind sea because it is dominated by the swell component. Such a condition is common in the trade wind regions with IWA value around 0.4 (light orange color). However, there are exceptions, such as the Gulf of Mexico, which has a large IWA of 0.65 due to the relatively short wind fetch in its semi-enclosed seas. The largest IWA values occur in the Coral Sea off east Australia and in the region northeast of Madagascar, which are the result of high wind events during the summer. Higher IWA values are also observed in the Caribbean Sean as well as in the Arabian Sea compared to open ocean along the same latitude. In Figure 3d, the visual representation of the spectral width is quite different from the other wave metrics. No evident disparity is observed between the storm track and the trade wind regions as the SWH and IWA do. The stand-out areas with larger spectral width correspond to the swell propagation route, typical in the Pacific Ocean and the Indian Ocean (Zheng et al., 2018). By comparing the spatial patterns in Figures 3a–3d, the Southern Ocean is worth a particular mention. The strong winds associated with the frequent occurrence of storms over this region generate high waves, which remain rather young in spite of the long fetch encountered by these waves. Their spectral width is intermediate between that of wind sea and swell. The underlying surface current of the Antarctic Circumpolar Current (ACC) may also impact this pattern. The ACC, known for its strong and persistent eastward flow, exerts a profound influence on the generation and propagation of ocean waves in the Southern Ocean region.

In addition to the mean state, variability of the wind-wave ensemble is analyzed through histograms of the four variables at two selected locations. As shown in Figure 4, in the Southern Hemisphere trade wind region (20°S,150°W; orange), U10 is less variable with a peak value of 7 $m \cdot s^{-1}$, compared to the Southern Ocean

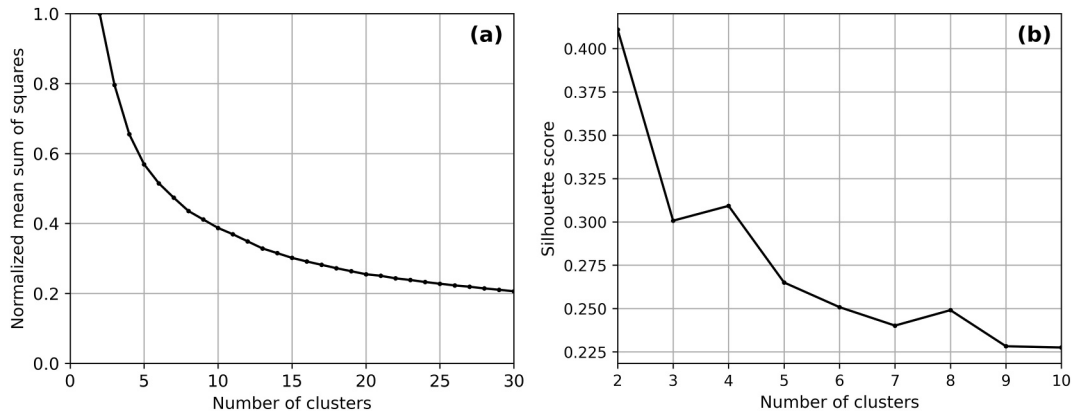

Figure 5. Performance evaluation of k-means clustering by (a) normalized mean sum of squared distance and (b) Silhouette score relative to the number of predefined clusters.

(50°S,150°W; blue) with a peak of 10 $m \cdot s^{-1}$. In the trade wind region where the swell is more prevalent, SWH has a smaller spread with values ranging from 1.5 m to 4 m compared to the wider distribution in the Southern Ocean in Figure 4b. This larger variability in the extratropics is due to the diversity of sea state development with IWA values ranging from 0 to 1.2 as in Figure 4c. The distributions of spectral width displayed in Figure 4d both follow a bell shape, in contrast to the other three variables. Compared with the region around 20°S, the region around 50°S located within the swell propagation route, exhibits a peak of distribution shifted toward smaller spectral width value possibly resulting from consistent wind forcing across the southern westerlies. These compiled wind-wave metrics provide a comprehensive characterization of sea state variability at the global scale.

3. Results of k-Means Clustering

3.1. The Clustering Method

K-means clustering is chosen here due to its overall simplicity in implementation (Lloyd, 1982) as well as its broad applications in wave related studies (Camus et al., 2014; Lobeto et al., 2022). The k-means clustering algorithm is a popular unsupervised learning method that groups a data set into a specified number of clusters based on similarities. Taking multivariate observational data vectors $\mathbf{X} = (x_1, x_2, \ldots, x_n)$ as an example, it is grouped into a priori k clusters each characterized by a corresponding centroid u_j as the prototype. The initial positions of centroids are randomly selected and then optimized through an iterative process. Each data point is assigned to a cluster by minimizing the point-to-centroid distance within each group. That is to say, the clusters are separated from each other in the feature space as each cluster captures a different pattern or group within the data. In this study, k-means clustering is implemented using the scikit learn (sklearn) module in Python. The 'KMeans' function with the option of init = 'k-means++' for distant centroid initialization, is used to ensure optimal performance.

The 4-dimensional wind-wave ensemble, consisting of U_{10}, SWH, IWA, and the spectral width ε_v, is used as input for the clustering. However, each dimension of the ensemble has a different scale, which can lead to unbalanced weights in the calculation of point-to-centroid distance. To address this issue, each dimension is normalized to the range of [0, 1] using a modified min-max normalization method. We take the 0.5 and 99.5 percentile of each variate as the lower and upper limits, rather than the minimum and maximum values, to avoid any impact caused by the extremes. Observations below the lower limit and above the upper limit are excluded from the clustering input. This normalization procedure has been demonstrated to be necessary for several oceanographic studies (Fairley et al., 2020; Sun et al., 2021).

In the k-means clustering algorithm, the optimal number of clusters is often determined using two methods: the elbow test and the Silhouette score (Figure 5). The elbow test displays the relationship between the normalized within-cluster sum of squares (the sum of the square distance between points in a cluster and the cluster centroid) and the number of clusters, with the optimal number of clusters being identified as the elbow point in the graph, as shown in Figure 5a. In this case, the elbow point is found to be $k = 7$; however, $k = 6$ was finally selected on

Table 1
Percentage and Centroids of the Clustered Sea States

Class index	Proportion [%]	U_{10} $[m \cdot s^{-1}]$	SWH [m]	IWA	ε_v
1	26.58	4.46	1.73	0.13	0.60
2	26.01	6.94	1.99	0.36	0.56
3	18.26	7.60	3.11	0.15	0.63
4	13.28	9.88	2.62	0.65	0.53
5	10.37	11.13	4.72	0.39	0.62
6	5.51	15.22	5.49	0.80	0.58

account of the geophysical status portrayed by the wind regimes (low, medium, and high) and dominance of wave systems (wind sea and swell). Note that only the regular sea state conditions are included and the exceptions, such as growing sea state are not considered here. The Silhouette score (Figure 5b), on the other hand, measures the separation distance between data points among different clusters and quantifies the goodness of clustering results. A score of one indicates a good separation between clusters and a score of −1 means that samples may have been assigned to the wrong cluster. As noticed, the Silhouette score does not significantly change as the number of clusters increases from 6 to 10, implying a stable clustering performance. This, along with the results of the wind-wave properties analysis at the global scale (presented in the following section), supports the cluster number choice of 6.

3.2. Characteristics of Wind-Wave Classes

The k-means clustering divides all the data points into 6 classes, each characterized by a unique centroid as listed in Table 1. The class index is ordered in terms of the wind speed magnitude. Class 1 has the lowest U_{10} centroid of 4.46 $m \cdot s^{-1}$, while class 6 corresponds to 15.22 $m \cdot s^{-1}$. As shown, classes 1 and 2 are the most populated with a comparable proportion, while class 6 is the least. This is in line with the global wind speed distribution that the low to moderate winds are the most common and high winds are relatively rare. Figure 6 illustrates the results of the k-means clustering by plotting wave metrics against U_{10}, and also shows the distribution of the classified groups. Each observation of [U_{10}, SWH, IWA, ε_v] is uniquely assigned to a group by the clustering algorithm, meaning that all data points are non-overlapping in the 4-dimensional feature space. However, the overlapping contour lines in Figure 6 suggest that the 6 clusters are not completely separated in either of the two-dimensional spaces. This highlights the benefits of classifying the sea state based on multivariate observations.

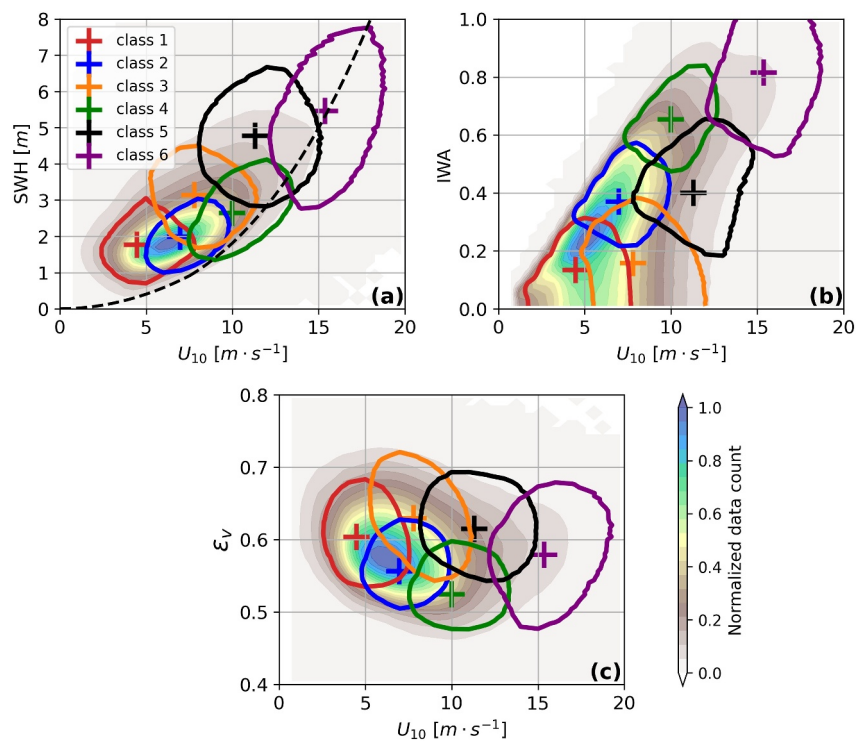

Figure 6. Clustering results depicted by the wind-wave relation of (a) SWH, (b) IWA and (c) the spectral width ε_v versus U_{10}. Color scales denote the data count at bins of 0.1 m, 0.03, 0.01 and 0.5 $m \cdot s^{-1}$ for SWH, IWA, ε_v and U_{10}, respectively. The dashed curve in (a) denotes the SWH-U_{10} relationship used in the WAM model to display the wind sea and swell dominance. The background color is the data count and the overlapped contours are 20% of the maximum data count for each cluster, with the cluster centroid marked by the plus sign in the same color.

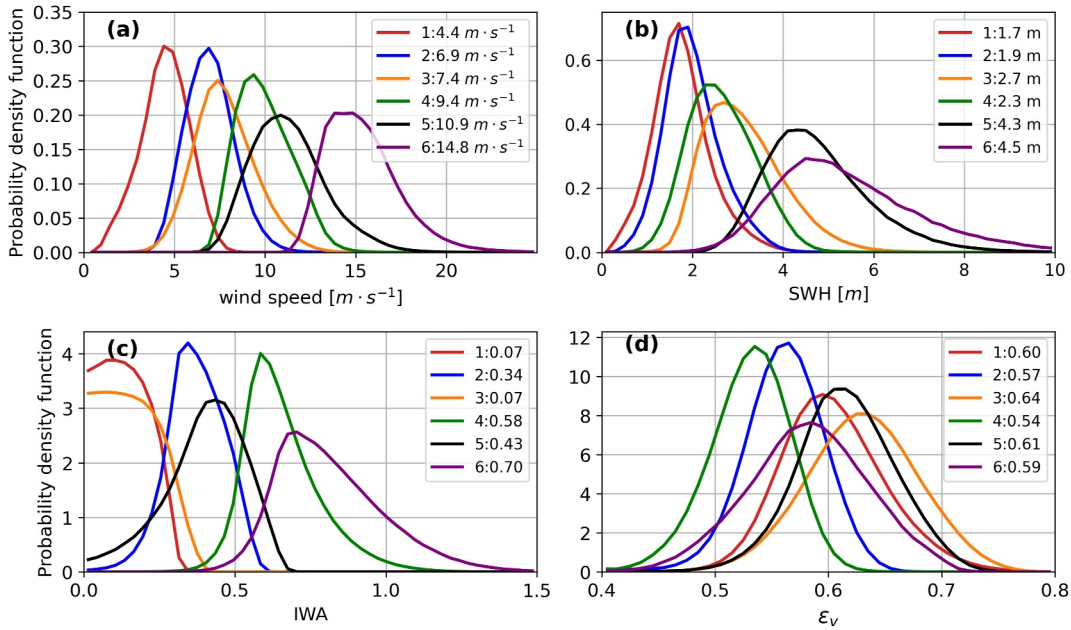

Figure 7. Probability density function (PDF) for each classified group of (a) U_{10}; (b) SWH and (c) IWA and (d) ε_v. Bins used to generate the PDFs are the same as Figure 6. The mode values of each PDF are annotated in the subplots.

SWH versus U_{10} in Figure 6a mostly lies above the wind-wave growth curve (dashed black), consistent with the previous findings (G. Chen et al., 2002; Jiang & Chen, 2013). The majority of the data points fall within the swell dominance regime and class 6 of high winds is distributed across the wind sea to swell dominance. It is observed that the variation of SWH centroids versus U_{10} increases can be divided into two groups, with the SWH centroids of classes 3 and 5 and 6 larger than that of classes 1 and 2 and 4. For example, class 3 has a lower wind speed compared to class 4 by 2.3 m $\cdot$ s^{-1}, but it has a higher SWH by 0.5 m due to its strong swell dominance with a low IWA centroid as depicted in Figure 6b. An opposite trend is observed in ε_v relative to U_{10} in Figure 6c. By comparing the classification results, the mixed sea state conditions could also be traced between classes. For instance, class 2 and 4 exhibit the lowest spectral width values, suggesting that they might correspond to the case of unimodal spectrum. While class 2 and class 5 have comparable IWA centroids, their spectral width values differ. These results provide a more detailed classification of swell dominance based on the maturity of the sea state, which complements the traditional index.

The variability of each class is presented through the probability density function (PDF) of four wind and wave variables in Figure 7. PDFs of U_{10} show a shift toward higher values from class 1 to class 6, with varying disparities in wind speed. While for SWH (Figure 7b), class 3 and class 4 almost overlap due to strong swell dominance in class 3, as reflected by the low IWA (Figure 7c). Classes 4 and 5 also have evident deviation, with a 1.4 m $\cdot$ s^{-1} difference in U_{10} and 1.8 m in SWH, reflecting a more pronounced sea state development of class 5 as indicated by the smaller IWA. Such disparity is not evident in the spectral width as class 1/3/5/6 are basically overlapped as in Figure 7d. It is worth noting that, based on the IWA criterion, only classes 4 and 6 contain underdeveloped wind sea, while all other classes correspond to swell. Within class 6, only 36.2% of the data points correspond to the developing wind sea (IWA > 0.84), which confirms their rarity in the global open ocean (Hanley et al., 2010). The relation of MSV and MSS versus wind speed is quite similar to that of SWH given their strong correlation with SWH as shown in Figure 2 (not shown for simplicity).

3.3. Global Occurrence of Wind-Wave Classes

Based on the 3 years data set of SWIM observations, the occurrence frequency of each wind-wave class is presented in Figure 8. Class 1, which represents low sea state combined with small IWA and medium to high ε_v, is predominantly found at the boundary between the trade winds and the westerlies. This condition occurs approximately 75% of the time in the Northwest Pacific Ocean and the North Indian Ocean, as well as off the California coast and the Gulf of Guinea in West Africa. In contrast, these conditions are rare in the Southern

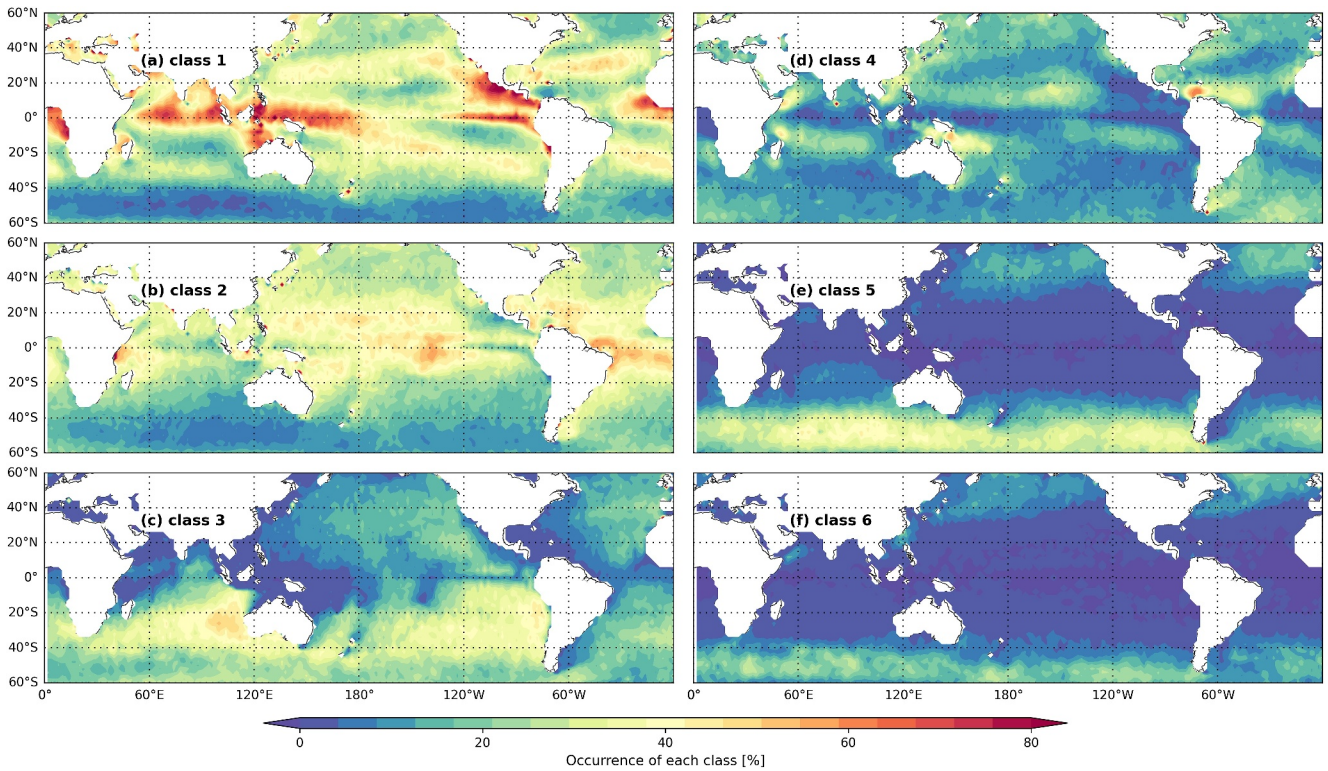

Figure 8. Global occurrence of frequency of each class at the latitude/longitude bin of 2°.

Hemisphere storm track, with less than 10% occurrence, compared to the 30% in the Northern Hemisphere storm track. This difference is attributed to the persistent low wind and SWH conditions during the summer season in the northern extratropics. Class 2, associated with moderate IWA and spectral width, is most commonly found in the trade winds, as seen in Figure 8b. This class has weaker swell contribution compared to class 1, indicated by its higher IWA value. It has a distinct global distribution across various ocean basins. In the Pacific Ocean and Atlantic Ocean, class 2 dominates between 25 °*N* and 25 °*S* with a peak occurrence around the equator. In the Indian Ocean, class 2 overlaps with class 1 but with a much lower occurrence. Storm tracks in both hemispheres have a low chance of experiencing this wave condition. On the other hand, class 3 with similar winds but lower IWA, is characterized by its unique pattern as shown in Figure 8c. It roughly traces the path of ocean swell from the southern storm track to the west coasts of Australia, Africa, and South America. This sea state is rare off the east coasts of the oceans, but it occurs about 20% of the time in the Southern Ocean, surpassing class 1 and class 2.

Class 4 of developed sea state with moderate to high winds, has a distinct global pattern, as shown in Figure 8d. It mainly fills in the gaps of class 1 in the trade wind regions, and two bands of high occurrence around 30% is observed in the Pacific Ocean, Atlantic Ocean, and Indian Ocean. The Gulf of Mexico also has frequent occurrence up to 60%, which may be due to the wave development in the semi-enclosed seas. Classes 5 and 6, which correspond to high sea states with similar SWH but different IWA (higher for class 6), have similar spatial features and are absent in the tropics, marked by dark blue in the figure. They both mainly occur close to storm generation, with a high presence in the Southern Hemisphere. Class 5 has a relatively broader distribution, extending to latitudes of 30° in both hemispheres, while class 6 with the youngest waves (largest IWA) is largely confined to latitudes beyond 40°. Class 6 is also found off the Southeast Asia coast and off the coast of Somalia, where storm-related high winds are occasional. It is worth mentioning the concept of mixed sea conditions from the spatial point of view. Class 2, mostly found in tropical regions, is more prone to unimodal spectra. Despite sharing the same wave age, class 5 primarily occurs in the Southern Ocean with a higher possibility of mixed sea state.

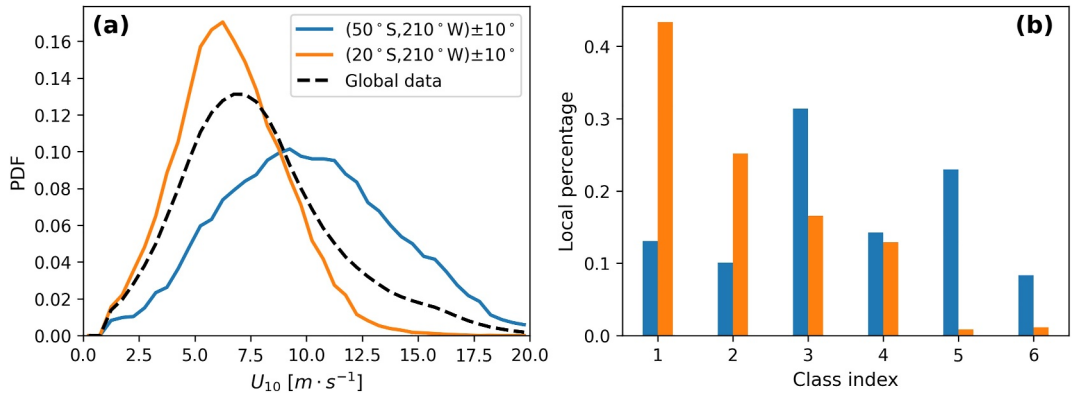

Figure 9. (a) An example of wind speed distribution over two distinct regions compared to the global distribution (black dashed curve). (b) The proportion of each class based on the clustering results with 4 input variables of the global data set. Blue and orange color in both plots denote the region 50°S ± 10° and 20°S ± 10° at longitude of 150°W ± 10°, respectively.

4. Discussion

In this study, the k-means clustering algorithm is implemented with 4 wind-wave variables (U_{10}, SWH, IWA and ε_v). Since the classification results might be influenced by the number of input variables, we also repeated the clustering process with 6 variables additionally including MSS and MSV. The clustering results obtained using 6 variables are quite similar to those obtained using 4 variables (not shown). This suggests that the additional two variables of MSS and MSV do not significantly contribute to the differentiation of data points within the clustering algorithm. We also compute the Adjusted Rand Index (ARI) to quantify the similarities between these clustering results. ARI compares how pairs of samples are assigned to the same or different clusters in the two clustering results by taking into account the agreement between clusters while considering chance (Hubert & Arabie, 1985). An ARI value close to 0 means random labeling of clusters and a higher value corresponds to better similarity between the clustered results with 1 to a perfect match. The ARI value computed between 4 and 6 input variables is 0.751, implying that 75% of the pairs of data points are consistently assigned to the same clusters across the two clustering results, after accounting for random chance. This further indicates that the exclusion of MSV and MSS does not significantly influence the classification results, which is probably because these two variables are in fact highly correlated to SWH (see Section 2.2).

Another important point is that the reported clustering results are obtained based only on 3 years of data from CFOSAT. Clusters are then conditional on the variability within that period and the representation in CFOSAT data. Clustering based on a different, possibly longer period, may result in differences in cluster characteristics due to temporal changes in variability, such as those driven by climate change, or due to the characteristics of the data itself, such as differences in numerical model representations. Similarly, clustering performed on regional data could also yield different results, possibly related to local impacts or specific environmental conditions. In this context, we conducted a sensitivity analysis within the available data set. As a demonstrative test, we applied clustering to a 1-year data set from 2022 and found that the resulting clusters were similar. This suggests stability of the variance within the 3-year record but a more detailed exploration of this issue is beyond the scope of this particular study.

Insufficient variability in the data space might result in clusters that are not representative of the full range of sea state conditions. Two regions of interest are selected as shown in Figure 9 to illustrate the potential impact of localized sea state conditions based on the 3-year CFOSAT data set. The wind speed PDF is presented in Figure 9a with the global wind speed PDF plotted for comparison. The region (20°S) in trade wind exhibits a greater chance of lower wind speed, while that in storm tracks (50°S) shifts toward higher values. This indicates that the localized regions may exhibit different diversity of sea state conditions. Assuming that one tries to establish the classification based on a limited number of regions, certain classes might be missed. The proportion of each clustered class over these two regions displayed in Figure 9b also evidences the need for the input data to be as global as possible. The region in trade wind (orange bar) is dominated by class 1 and 2 of low winds. This contrasts to the dominant position of class 3 and 5 in the storm tracks (blue bar). In addition, we also looked into the wind wave parameters at the NDBC buoy locations (not shown) and came to a similar conclusion that

localized regions are not able to capture the full spectrum of sea state variability. The k-means clustering algorithm relies on the distribution of data points to determine the cluster centroids. Insufficient variability may lead to centroids that are biased toward the predominant conditions in the local data set, potentially neglecting important but less frequent sea state patterns.

In addition, seasonality might also influence the clustering results, particularly in the Northern Hemisphere mid-latitudes over a relatively short temporal period. The winter storms and increased wind speeds during colder months, result in distinct clustering patterns compared to the calmer conditions observed in summer. This seasonality effect can create clusters that are subject to the conditions prevalent during the period of interest if it is shorter than 1 year. To summarize, the temporal period of the input data set may not significantly influence the clustering results when the data set provides, to a maximum extent, a global coverage of sea state conditions longer than 1 year. In the context of wind-wave classification, the key is mostly on capturing the spatial patterns and variability of sea state parameters across different regions. A data set spanning multiple years or even decades might be better regarding its representativeness of long-term climate trends; however, the primary emphasis lies in adequate spatial coverage to effectively characterize global wind-wave conditions.

5. Summary and Perspectives

A wind-wave ensemble consisting of wind speed, significant wave height, inverse wave age, and the spectral width, mean square slope and mean surface orbital velocity is created based on concurrent scatterometer and SWIM observations from the CFOSAT satellite. The accuracy of wind speed and wave height estimations is verified using *in situ* buoy data, with satisfactory results. Both MSS and MSV are excluded from the clustering input given their strong correlation with SWH. The k-means clustering using the 4-dimensional wind-wave metrics complements previous wind-wave classification schemes by incorporating a multivariate description of sea state conditions.

The resulting spatial representation of the 6 clustered classes clearly shows the overall sea state variability at both global and regional scales. Classes 2 and 3, with different SWH but comparable wind speeds, distinguish the degree of swell dominance through the distinct IWA and spectral width. The new classification refines the regional variations of swell-dominant areas previously identified in (G. Chen et al., 2002). Swell waves (classes 1–5) are also confirmed to be present across most of the global ocean, even in the storm tracks. Class 5, characterized by high winds and low IWA, is predominantly positioned across the extratropics, with an occurrence of about 35% in the southern storm track (refer to Figure 8e). On the other hand, class 6, with higher U_{10} and larger IWA, only occurs 20% of the time as documented in Hanley et al. (2010), further illustrating the rare presence of pure wind seas and their prevalence in the storm tracks.

The impact of sea state on friction velocity, a key factor in upper ocean dynamics, has been well documented (Donelan et al., 1993; S. Li et al., 2020). A recent study has shown that at low winds, the friction velocity increases in the presence of swell, based on a set of *in situ* wind and wave observations (Vincent et al., 2020). Wind stress over regions where classes 1, 2, and 3 are located may therefore deviate from the pure wind sea conditions due to swell impact. The deviation degree might be related to the level of swell dominance and/or the stage of sea state development. As suggested by Vincent et al. (2020); S. Chen et al. (2020), an empirical algorithm that relates the drag coefficient to the local sea state may be a practical approach to accounting for the impact of swell. In addition, the influence of underlying surface current might also be noted, particularly in the Southern Ocean. The almost permanently strong winds combined with the absence of continent lead to the formation of wind-driven waves that can develop and remain under the action of wind over long fetch during their propagation. These extended fetch as well as the wave-current interaction might jointly contribute to the generation of large waves with enhanced energy content. Consequently, the friction velocities at the ocean surface are intensified, facilitating increased momentum and energy exchange between the atmosphere and the ocean (S. Chen et al., 2020).

The classification presented in this study provides a more detailed description of sea state conditions. Unlike a traditional binary classification, which only distinguishes between wind-wave and swell dominance, the six-class classification scheme allows for the identification of various intermediate states and transitional conditions. It shall help identify the subtle trend and patterns in wave systems characterized by each class. Compared to the spatial coverage of SWIM that extends over approximately 70 km × 90 km, spaceborne synthetic aperture radar images typically provide higher spatial resolution at around 20 km × 20 km, with the potential to capture details at even finer scales of 5–10 km. By integrating data from both instruments, we could possibly gain insights into

wave characteristics across various scales, from mesoscale current features including eddies and frontal boundaries to finer-scale phenomena such as wave-ice interactions. Further investigations that combine satellite observations and *in situ* measurements shall also help understand and compare the momentum transfer under different wind-wave interaction cases.

Building upon the advantages of obtaining wind and wave estimates from CFOSAT mission, several aspects shall be further explored. On one hand, such a wind-wave classification scheme might be included in future satellite processing chains to account for sea-state conditions in ocean surface parameter retrieval. For example, better characterizing wind and wave conditions is crucial for interpreting Doppler shift retrievals from synthetic aperture radar, Doppler-based scatterometers and future satellite missions. These retrieval methods rely on capturing the Doppler shift caused by various surface motions, including wave orbital motion, wind-induced surface drift, and underlying surface currents. By classifying wind and wave conditions using methods such as k-means clustering, we can better understand the individual contributions of these factors to the Doppler shift signal. In addition, while the potential for collocation and co-variability analyses between on-orbit missions such as CFOSAT, Sentinel-1, and SWOT may enhance our sea state classification scheme, it is important to recognize the limitations and differences in acquisition methods across these missions. SWOT, primarily designed for topographic missions, carries both the Poseidon altimeter and the KARIN interferometer, but the latter is not yet operational for wave height retrievals which makes its contribution to sea state classification at present is limited. As more data becomes available, the simultaneous measurements of sea surface height and significant wave height might provide new sights into this direction. On the other hand, applying the k-means clustering to long-term wave reanalysis data, like ERA5 presents a new opportunity to look into wave climate in both global and local framework. The consistency of such wave clusters and their patterns across the long-term data set, along with the temporal change of sea state clusters, deserves to be examined and will constitute our follow-on study. Based on which, identification of distinct wave regimes such as storm waves, swell waves, and possibly extreme waves, is possible for a comprehensive understanding of wave climate variability. Moreover, by examining the temporal evolution and spatial distribution of these wave clusters, the regional wave climate trends including changes in wave height, directionality, and frequency over extended periods could be investigated. The integration of such clustering with wave reanalysis data not only enhances our ability to characterize wave climates but also contributes to improved wave forecasting, coastal management, and relevant decision-making.

Data Availability Statement

All CFOSAT products are provided by the courtesy of CNSA and CNES. The CFOSAT Level-2 products are publicly accessible via https://www.aviso.altimetry.fr/en/missions/current-missions/cfosat/access-to-data.html. The buoy data of wind speed and significant wave height are provided by NDBC and publicly available via http://www.marineinsitu.eu/. The collocated data set archiving is publicly available on Researchgate via (H. Li, 2024).

Acknowledgments
This study is supported by the National Science Foundation of China under grant of 42476182 and 42027805, the Centre National d'études Spatiales (CNES) under the TOSCA program. Work by D. Vandemark was performed and funded by award ST13301CQ0050/1332KP22FNEED0042 provided by NOAA-NESDIS. We would like to express our gratitude to the three anonymous reviewers for their constructive comments and suggestions that greatly helped us to improve the manuscript.

References

Aouf, L., Hauser, D., Chapron, B., Toffoli, A., Tourain, C., & Peureux, C. (2021). New directional wave satellite observations: Towards improved wave forecasts and climate description in southern ocean. *Geophysical Research Letters*, *48*(5), e2020GL091187. https://doi.org/10.1029/2020GL091187

Ardhuin, F., Chapron, B., & Collard, F. (2009). Observation of swell dissipation across oceans. *Geophysical Research Letters*, *36*(6), 1–5. https://doi.org/10.1029/2008GL037030

Ardhuin, F., Rogers, E., Babanin, A. V., Filipot, J.-F., Magne, R., Roland, A., et al. (2010). Semiempirical dissipation source functions for ocean waves. Part I: Definition, calibration, and validation. *Journal of Physical Oceanography*, *40*(9), 1917–1941. https://doi.org/10.1175/2010JPO4324.1

Camus, P., Menéndez, M., Méndez, F. J., Izaguirre, C., Espejo, A., Cánovas, V., et al. (2014). A weather-type statistical downscaling framework for ocean wave climate. *Journal of Geophysical Research: Oceans*, *119*(11), 7389–7405. https://doi.org/10.1002/2014JC010141

Cavaleri, L., Fox-Kemper, B., & Hemer, M. (2012). Wind waves in the coupled climate system. *Bulletin of the American Meteorological Society*, *93*(11), 1651–1661. https://doi.org/10.1175/BAMS-D-11-00170.1

Chen, G., Chapron, B., Ezraty, R., & Vandemark, D. (2002). A global view of swell and wind sea climate in the ocean by satellite altimeter and scatterometer. *Journal of Atmospheric and Oceanic Technology*, *19*(11), 1849–1859. https://doi.org/10.1175/1520-0426(2002)019⟨1849:AGVOSA⟩2.0.CO;2

Chen, S., Rutgersson, A., Yin, X., Xu, Y., & Qiao, F. (2020). On the first observed wave-induced stress over the global ocean. *Journal of Geophysical Research: Oceans*, *125*(12), e2020JC016623. https://doi.org/10.1029/2020JC016623

Collard, F., Ardhuin, F., & Chapron, B. (2009). Monitoring and analysis of ocean swell fields from space: New methods for routine observations. *Journal of Geophysical Research*, *114*(C7), 1–15. https://doi.org/10.1029/2008JC005215

Desbiolles, F., Bentamy, A., Blanke, B., Roy, C., Mestas-Nuñez, A. M., Grodsky, S. A., et al. (2017). Two decades [1992–2012] of surface wind analyses based on satellite scatterometer observations. *Journal of Marine Systems*, *168*, 38–56. https://doi.org/10.1016/j.jmarsys.2017.01.003

Donelan, M. A., Babanin, A. V., Young, I. R., & Banner, M. L. (2006). Wave-follower field measurements of the wind-input spectral function. Part II: Parameterization of the wind input. *Journal of Physical Oceanography*, *36*(8), 1672–1689. https://doi.org/10.1175/JPO2933.1

Donelan, M. A., Dobson, F. W., Smith, S. D., & Anderson, R. J. (1993). On the dependence of sea surface roughness on wave development. *Journal of Physical Oceanography*, *23*(9), 2143–2149. https://doi.org/10.1175/1520-0485(1993)023⟨2143:OTDOSS⟩2.0.CO;2

Donelan, M. A., Drennan, W. M., & Katsaros, K. B. (1997). The air–sea momentum flux in conditions of wind sea and swell. *Journal of Physical Oceanography*, *27*(10), 2087–2099. https://doi.org/10.1175/1520-0485(1997)027⟨2087:TASMFI⟩2.0.CO;2

Ebuchi, N., Graber, H. C., & Caruso, M. J. (2002). Evaluation of wind vectors observed by quikscat/seawinds using ocean buoy data. *Journal of Atmospheric and Oceanic Technology*, *19*(12), 2049–2062. https://doi.org/10.1175/1520-0426(2002)019⟨2049:EOWVOB⟩2.0.CO;2

Elfouhaily, T., Chapron, B., Katsaros, K., & Vandemark, D. (1997). A unified directional spectrum for long and short wind-driven waves. *Journal of Geophysical Research*, *102*(C7), 15781–15796. https://doi.org/10.1029/97JC00467

Fairley, I., Lewis, M., Robertson, B., Hemer, M., Masters, I., Horrillo-Caraballo, J., et al. (2020). A classification system for global wave energy resources based on multivariate clustering. *Applied Energy*, *262*, 114515. https://doi.org/10.1016/j.apenergy.2020.114515

Gao, Y., Schmitt, F. G., Hu, J., & Huang, Y. (2023). Probability-based wind-wave relation. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.1085340

Grachev, A. A., & Fairall, C. W. (2001). Upward momentum transfer in the marine boundary layer. *Journal of Physical Oceanography*, *31*(7), 1698–1711. https://doi.org/10.1175/1520-0485(2001)031⟨1698:UMTITM⟩2.0.CO;2

Hanley, K. E., Belcher, S. E., & Sullivan, P. P. (2010). A global climatology of wind–wave interaction. *Journal of Physical Oceanography*, *40*(6), 1263–1282. https://doi.org/10.1175/2010JPO4377.1

Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., et al. (1973). Measurements of wind-wave growth and swell decay during the joint north sea wave project (Jonswap). *Ergnzungsheft zur Deutschen Hydrographischen Zeitschrift Reihe*, *8*, 1–95.

Hauser, D., Soussi, E., Thouvenot, E., & Rey, L. (2001). Swimsat: A real-aperture radar to measure directional spectra of ocean waves from space—Main characteristics and performance simulation. *Journal of Atmospheric and Oceanic Technology*, *18*(3), 421–437. https://doi.org/10.1175/1520-0426(2001)018⟨0421:SARART⟩2.0.CO;2

Hauser, D., Tourain, C., Hermozo, L., Alraddawi, D., Aouf, L., Chapron, B., et al. (2021). New observations from the swim radar on-board cfosat: Instrument validation and ocean wave measurement assessment. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(1), 5–26. https://doi.org/10.1109/TGRS.2020.2994372

Hell, M. C., Ayet, A., & Chapron, B. (2021). Swell generation under extra-tropical storms. *Journal of Geophysical Research: Oceans*, *126*(9), e2021JC017637. https://doi.org/10.1029/2021JC017637

Högström, U., Smedman, A., Semedo, A., & Rutgersson, A. (2011). Comments on "a global climatology of wind–wave interaction". *Journal of Physical Oceanography*, *41*(9), 1811–1813. https://doi.org/10.1175/jpo-d-10-05015.1

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218. https://doi.org/10.1007/bf01908075

Janssen, P. A. E. M., Saetra, Ø., Bidlot, J.-R., & Beljaars, A. (2003). Impact of sea state on atmosphere and ocean. In *Ecmwf Workshop on the Role of the Upper Ocean in Medium and Extended Range Forecasting, 13-15 November 2002* (pp. 143–157). ECMWF.

Jiang, H., & Chen, G. (2013). A global view on the swell and wind sea climate by the Jason-1 mission: A revisit. *Journal of Atmospheric and Oceanic Technology*, *30*(8), 1833–1841. https://doi.org/10.1175/JTECH-D-12-00180.1

Kinsman, B. (1965). *Wind waves: Their generation and propagation on the ocean surface*. Dover Publications, INC.

Kudryavtsev, V., Yurovskaya, M., & Chapron, B. (2021). Self-similarity of surface wave developments under tropical cyclones. *Journal of Geophysical Research: Oceans*, *126*(4), e2020JC016916. https://doi.org/10.1029/2020JC016916

Le Merle, E., Hauser, D., Peureux, C., Aouf, L., Schippers, P., Dufour, C., & Dalphinet, A. (2021). Directional and frequency spread of surface ocean waves from swim measurements. *Journal of Geophysical Research: Oceans*, *126*(7), e2021JC017220. https://doi.org/10.1029/2021JC017220

Li, H. (2024). [Dataset]. Wind wave colloc. https://doi.org/10.13140/RG.2.2.15718.64329

Li, S., Zou, Z., Zhao, D., & Hou, Y. (2020). On the wave state dependence of the sea surface roughness at moderate wind speeds under mixed wave conditions. *Journal of Physical Oceanography*, *50*(11), 3295–3307. https://doi.org/10.1175/JPO-D-20-0102.1

Li, X.-M. (2016). A new insight from space into swell propagation and crossing in the global oceans. *Geophysical Research Letters*, *43*(10), 5202–5209. https://doi.org/10.1002/2016GL068702

Liu, J., Lin, W., Dong, X., Lang, S., Yun, R., Zhu, D., et al. (2020). First results from the rotating fan beam scatterometer onboard cfosat. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(12), 8793–8806. https://doi.org/10.1109/TGRS.2020.2990708

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Lobeto, H., Menendez, M., Losada, I. J., & Hemer, M. (2022). The effect of climate change on wind-wave directional spectra. *Global and Planetary Change*, *213*, 103820. https://doi.org/10.1016/j.gloplacha.2022.103820

Longuet-Higgins, M. S. (1975). On the joint distribution of the periods and amplitudes of sea waves. *Journal of Geophysical Research (1896-1977)*, *80*(18), 2688–2694. https://doi.org/10.1029/JC080i018p02688

Miles, J. W. (1957). On the generation of surface waves by shear flows. *Journal of Fluid Mechanics*, *3*(2), 185–204. https://doi.org/10.1017/S0022112057000567

Paulson, C. A. (1970). The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. *Journal of Applied Meteorology and Climatology*, *9*(6), 857–861. https://doi.org/10.1175/1520-0450(1970)009⟨0857:TMROWS⟩2.0.CO;2

Risien, C. M., & Chelton, D. B. (2008). A global climatology of surface wind and wind stress fields from eight years of quikscat scatterometer data. *Journal of Physical Oceanography*, *38*(11), 2379–2413. https://doi.org/10.1175/2008JPO3881.1

Saulnier, J.-B., Clément, A., De, O., Falcão, A. F., Pontes, T., Prevosto, M., & Ricci, P. (2011). Wave groupiness and spectral bandwidth as relevant parameters for the performance assessment of wave energy converters. *Ocean Engineering*, *38*(1), 130–147. https://doi.org/10.1016/j.oceaneng.2010.10.002

Semedo, A., Sušelj, K., Rutgersson, A., & Sterl, A. (2011). A global view on the wind sea and swell climate and variability from era-40. *Journal of Climate*, *24*(5), 1461–1479. https://doi.org/10.1175/2010JCLI3718.1

Shimura, T., Hemer, M., Lenton, A., Chamberlain, M. A., & Monselesan, D. (2020). Impacts of ocean wave-dependent momentum flux on global ocean climate. *Geophysical Research Letters*, *47*(20), e2020GL089296. https://doi.org/10.1029/2020GL089296

Snodgrass, F. E., Hasselmann, K. F., Miller, G. R., Munk, W. H., Powers, W. H., & Deacon, G. E. R. (1966). Propagation of ocean swell across the pacific. *Philosophical Transactions of the Royal Society of London - Series A: Mathematical and Physical Sciences*, *259*(1103), 431–497. https://doi.org/10.1098/rsta.1966.0022

Stopa, J. E., Ardhuin, F., Chapron, B., & Collard, F. (2015). Estimating wave orbital velocity through the azimuth cutoff from space-borne satellites. *Journal of Geophysical Research: Oceans*, *120*(11), 7616–7634. https://doi.org/10.1002/2015JC011275

Sun, Q., Little, C. M., Barthel, A. M., & Padman, L. (2021). A clustering-based approach to ocean model–data comparison around Antarctica. *Ocean Science*, *17*(1), 131–145. https://doi.org/10.5194/os-17-131-2021

Vincent, C. L., Graber, H. C., & Collins, I. C. O. (2020). Effect of swell on wind stress for light to moderate winds. *Journal of the Atmospheric Sciences*, *77*(11), 3759–3768. https://doi.org/10.1175/JAS-D-19-0338.1

Xu, Y., Hauser, D., Liu, J., Si, J., Yan, C., Chen, S., et al. (2022). Statistical comparison of ocean wave directional spectra derived from swim/cfosat satellite observations and from buoy observations. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–20. https://doi.org/10.1109/TGRS.2022.3199393

Yang, J., & Zhang, J. (2019). Accuracy assessment of hy-2a scatterometer wind measurements during 2011–2017 by comparison with buoys, ascat, and era-interim data. *IEEE Geoscience and Remote Sensing Letters*, *16*(5), 727–731. https://doi.org/10.1109/LGRS.2018.2879909

Young, I. R., Zieger, S., & Babanin, A. V. (2011). Global trends in wind speed and wave height. *Science*, *332*(6028), 451–455. https://doi.org/10.1126/science.1197219

Zheng, C. W., Li, C. Y., & Pan, J. (2018). Propagation route and speed of swell in the Indian Ocean. *Journal of Geophysical Research: Oceans*, *123*(1), 8–21. https://doi.org/10.1002/2016JC012585