 doi:10.20944/preprints202001.0200.v1

Title

A global sea state dataset from spaceborne synthetic aperture radar wave mode data

Authors

Xiao-Ming Li[1], BingQing Huang[1,2]

Affiliations

1. Aerospace Information Research Institution, Chinese Academy of Sciences
2. University of Chinese Academy of Sciences

corresponding author: Xiao-Ming Li (lixm@aircas.ac.cn; lixm@radi.ac.cn)

Abstract

This dataset consists of integral ocean wave parameters of significant wave height (SWH) and mean wave period (MWP) data derived from the Advanced Synthetic Aperture Radar (ASAR) on board the ENVISAT satellite over its full life cycle (2002-2012) covering the global ocean. Both parameters are calibrated and validated against buoy data. A cross-validation between the ASAR SWH and radar altimeter (RA) data is also performed to ensure that the SAR-derived wave height data are of the same quality as the RA data. These data are stored in the standard NetCDF format, which are produced for each ASAR wave mode Level1B data provided by the European Space Agency. This is for the first time that a full sea state product is derived from spaceborne SAR data over the global ocean for a decadal temporal scale.

Background & Summary

The sea state is one of the key parameters of the "essential climate variables" (ECVs) defined by the Global Climate Observing System (GCOS) to meet the requirements of the climate change community. Spaceborne radar measurements of the sea state in terms of the significant wave height (SWH) and mean wave period (MWP), particularly observations from radar altimeters (RAs), have been available for a few decades[1]. Long-term RA measurements can reflect some wave height trends in the global oceans, and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

these trends might be associated with climate change[2]. Another radar sensor capable of measuring the sea state is known as spaceborne Synthetic aperture radar (SAR), which became available at the same time as RAs; consequently, both instruments were on board the Seasat[3] satellite launched in 1978. However, unlike nadir-looking RAs, SAR is a side-looking radar, which allows SAR to image large surface areas. Additionally, SAR can achieve a high spatial resolution in the azimuth (flight) direction through the "aperture synthetizing" technique[4]. In principle, spaceborne SAR should be able to effectively measure the sea state from space, as this technology images sea surface waves in two dimensions[5], at a high spatial resolution. However, as surface waves are in motion during the SAR imaging time (i.e., water particles are moving either toward or away from the radar system), the high-frequency components of ocean waves are missed (the "cut-off effect"), and the distortion of the spectrum during the imaging process of SAR[6-7]. Therefore, the SAR imaging of surface gravity waves is generally considered a nonlinear process[8], complicating the retrieval of ocean wave parameters from SAR data. Two-dimensional wave spectra predicted by ocean wave modeling (e.g., WAM[9]) or derived from other sources[10] must commonly be used as a priori (also called the "first guess") in the retrieval[11] to compensate for the lost and distorted ocean wave information during the SAR imaging. However, as a result of this compensatory approach, the retrieval of ocean wave parameters from SAR data has to rely on a priori information, which significantly limits SAR as an independent remote sensing instrument that can measure the sea state.

The wave mode (WM), which is dedicated to measurements of ocean wave, is a unique imaging mode of SAR. Although the WM covers a relatively area of the sea surface (approximately 6 km by 10 km), these data are automatically acquired by spaceborne SAR over the global oceans. From the European Remote Sensing Satellite missions (ERS-1, 1991 – 2000 and ERS-2, 1995 - 2011)[12-15] to the Environment Satellite

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

(ENVISAT) mission (2002 – 2012)[16-17] and the current Sentinel-1A/1B (2014 -)[18] and Chinese Gaofen-3 (2016 -) missions[19], WM data have been available for nearly 30 years and will continue to be acquired into the future, constituting a valuable dataset for global sea state measurements. On the basis of SAR WM data, some interesting investigations of global ocean waves, particularly with respect to the characteristics of ocean swells[20-22], have been reported. Such analyses can be performed because ocean swells are generally considered linearly or quasi-linearly imaged by SAR; thus, the abovementioned nonlinear inversion process can be "degraded" to a quasilinear approach[23-24], in which case a priori information is no longer needed. However, such a quasilinear inversion cannot yield full sea state parameters, e.g., the total SWH and MWP of both windsea and swell. Therefore, to retrieve sea state parameters from these valuable and global WM datasets, various parametric models that directly relate SAR-measured sea surface radar backscatter (radar cross section) to the sea state parameters of SWH and MWP have been proposed[14, 17-18]. Such algorithms can yield full sea state parameters without needing a priori information and can provide independent SAR measurements of global ocean waves. Here, we developed a global sea state dataset from the ENVISAT/Advanced Synthetic Aperture radar (ASAR) WM data acquired from 2002 to 2012 based on a parametric model[17], which we hereafter call the "CWAVE_ENV" model. This is for the first time that a global ocean dataset of full sea state parameters in a decadal temporal scale becomes publicly available based on spaceborne SAR data and we believe that this dataset, in conjunction with RA datasets that widely exploited at present, is valuable for global observations of ocean waves.

Methods

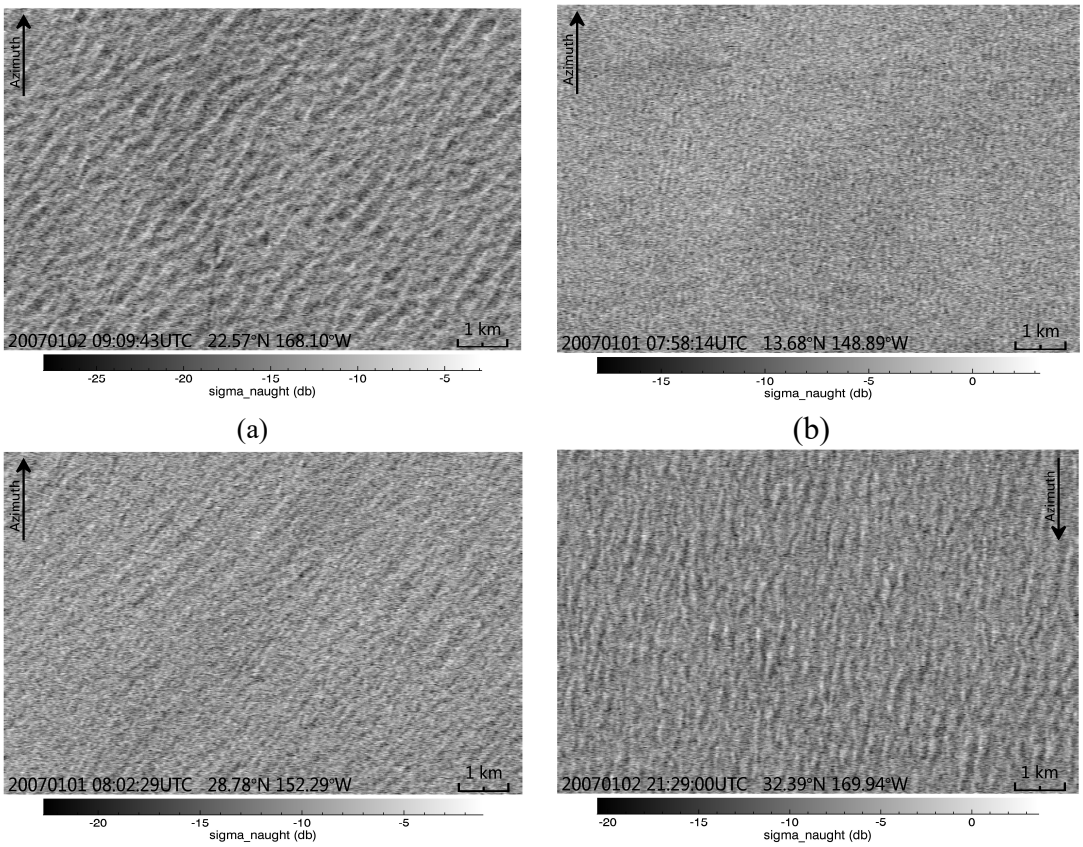
ASAR WM data

In the ERS-1 and ERS-2 missions, the SAR WM data were publicly available in the formation of two-dimensional image spectra in discrete formats, i.e. allocating image

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

spectrum energy in numbers of directional and frequency bins[25]. Beginning with the ENVISAT mission, the ASAR WM data in complex format[26] (i.e., consisting of a real part R_e and an imaginary part I_m) were provided to users; these data record both the magnitude and phase of the returned radar signals. The SAR image intensity (I) is therefore calculated as $I = R_e^2 + I_m^2$. By performing a radiometric calibration of intensity data, the normalized radar cross section, denoted σ_0, can be obtained and then used to retrieve sea state parameters.

ASAR WM data have a spatial coverage ranging from 6 km x 5 km to 10 km x 5 km over the sea surface. The distance between two consecutive acquisitions of WM data is 100 km. Examples of ASAR WM data acquired over the ocean are shown in Figure 1, which clearly displays patterns of ocean waves (swells).



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

Fig. 1 Examples of ASAR WM images acquired over the ocean. The acquisition date and location are marked on each image.

The parametric model "CWAVE_ENV" was applied to the ASAR WM data to generate the sea state parameters of SWH (h_s) and MWP (t_{m0}). This type of parametric model was first proposed for the reprocessed ERS-2 WM data[14]; the name "CWAVE" indicates the use of a C-band (SAR) wave retrieval algorithm, such as the widely used C-band geophysical model function "CMOD"[27], to retrieve sea surface wind fields from scatterometer and SAR data. Because the development and validation of parametric models have been described in detail in previous studies[14,17], only the rationale for using a parametric model is discussed here.

Although imaging mechanisms of ocean surface gravity waves by SAR remain further investigations, the measured radar backscatter from the sea surface is closely related to various sea state parameters (denoted W) through relations with a set of parameters (expressed as a vector, $\boldsymbol{S}(s_1, \dots, s_{ns})$). These parameters can be directly derived from SAR data, as expressed in equation 1.

$$W = a_0 + \sum_{1 \le i \le n_s} a_i s_i + \sum_{1 \le i \le j \le n_s} a_{i,j} s_i s_j \qquad (1)$$

In the above model, the sea state parameter W is expressed as linear combinations of ASAR image parameters $\boldsymbol{S}(s_1, \dots, s_{ns})$ with the extended coefficient vector $\boldsymbol{A}(a_0, \dots, a_{ns}, a_{11}, \dots, a_{n_s} a_{n_s})$. To also include nonlinearities as well as possible coupling among different parameters, a quadratic term is added to the equation (the third term in the equation). After the coefficient vector is determined, one can derive sea state parameters directly from the SAR-measured radar backscatter information of the sea surface.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

The least-square minimization approach is used to determine the coefficient vector $\boldsymbol{A}$ in equation 21 as defined in equation 2, where $(w^{(1)}, \boldsymbol{S}^{(1)}), \ldots, (w^{(N)}, \ldots, \boldsymbol{S}^{(N)})$ represents the available data pairs of SAR image parameters and the collocated tuning dataset of the integral wave parameter (SWH or MWP).

$$J_{cost}(A) = \sum_{j=1}^{N} (w^{(j)} - \sum_{i=0}^{n_A - 1} \boldsymbol{A}_i \boldsymbol{S}_i^j)^2 \quad (2)$$

After the coefficient vector is determined, one can derive the SWH or MWP directly from ASAR WM data using equation 1. The preliminary validation of the ASAR-derived h_s values using the CWAVE_ENV algorithm was conducted for a two-month (January and February 2017) dataset. Comparisons with the National Data Buoy Center (NDBC) *in situ* buoy measurements yielded a bias of 0.06 m and a root-mean-square-root (RMSE) of 0.70 m[17]. Here, we applied this parametric model to the entire dataset of the ASAR WM data of its full life cycle.

The entire ENVISAT mission ranged from March 2002 to April 2012. The ASAR data that we received from the European Space Agency (ESA) cover the period from December 2002 to April 2012. During the lifetime of ENVISAT, the ASAR instrument acquired WM data in vertical-vertical (VV) polarization with an incidence angle of 23º, except during two experimental periods, in which the acquired WM data had an incidence angle of approximately 33º. The first period ranged from January 24th to February 6th, 2007, and the second one ranged from March 6th to March 13th, 2007. From January 24th to January 30th, 2007, the WM data were acquired in horizontal-horizontal (HH) polarization. In addition to excluding the WM data acquired during these two experimental periods, the following criteria were applied to further screen the data.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

(i) The ASAR WM data acquired in polar regions were excluded from further processing because they might be affected by sea ice; thus, only the data acquired between 65ºS and 70ºN were used to generate sea state parameters.

(ii) Although ASAR WM images have a relatively small spatial coverage compared with images acquired in other modes, e.g., imaging mode and wide swath mode, the WM images are also affected by other sea surface features not related with ocean waves, e.g., oil spill, atmospheric features, and bright targets. To select only ASAR WM images that display a homogeneous sea surface (e.g., the cases shown in Figure 1) and derive sea state parameters, some parameters were used for automatic detection. We previously used the "homogeneity factor"[28] to classify ASAR WM images into homogenous and inhomogeneous classes; if the sea surface is "purely" homogeneous, this factor is equal to 1. Through the visual inspection of large amounts of both ERS-2/SAR and ENVISAT/ASAR WM data, the homogeneity factor was set to 1.05 as a threshold for selecting appropriate SAR WM data for retrieval. Approximately 94.42% of the data had a homogeneity factor lower than 1.05. After the aforementioned preprocessing steps, approximately 6.48 million ASAR WM data eventually were used to generate global ocean wave parameters.

In situ buoy data

In situ buoy measurements of sea state parameters were used to validate and calibrate the retrieved SWH and MWP based on the ASAR WM data. The GlobWave project (http://globwave.ifremer.fr/) collected a large amount of in situ buoy data from several buoy networks, including the Coastal Data Information Program (CDIP), the Center d'Etudes Techniques Maritimes Et Fluviales (CETMEF), the CORIOLIS, The European Center for Medium-Range Weather Forecasts (ECMWF), the Marine Environmental Data Section (MEDS) and the NDBC. We compared the different buoy

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

datasets collected in the GlobWave data portal and found that the dataset provided by the ECMWF contains more data (from 649 buoys collected between 2002 and 2012), than any of the other datasets. Therefore, we selected the ECMWF-provided buoy data (hereafter referred to as "ECMWF buoy data") for the evaluation and calibration of the ASAR-derived SWH.

However, the "mean wave period" in the ECMWF buoy dataset is the averaged wave period during a certain recording period, whereas the ASAR-retrieved MWP is the zero-upcrossing period (T_{m02}), as defined in equation 3. Therefore, we used the NDBC two-dimensional buoy spectrum (also accessed from the GlobWave data portal, hereafter called "NDBC buoy data") to calculate T_{m02} for comparison with the ASAR-retrieved MWP. The quality flag in the GlobWave NDBC buoy dataset is named *spectral_wave_density_qc_level*. The values of this flag are 0, 1, 2, 3 and 4, which represent *unknown*, *unprocessed*, *bad*, *suspect* and *good*, respectively. We used only good data for calibration and validation.

$$T_{m02} = \sqrt{m_0/m_2} \tag{3}$$

$$m_n = \sum_i f_i^n S_i \Delta f_i \tag{4}$$

In the above equations, where m_n is the n^{th} spectral moment, f_i is the i^{th} discrete frequency, Δf_i is the width of the i^{th} discrete frequency and S_i is the spectral density over the i^{th} frequency.

Both the ECMWF and NDBC buoy data were collocated with the ASAR WM data following the criteria that the temporal difference is less than 30 minutes and the spatial distance is less than 100 km. For cases in which several buoys satisfied the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

collocation criteria, only the measurements from the buoy nearest the corresponding ASAR WM data point were used for the validation and calibration.

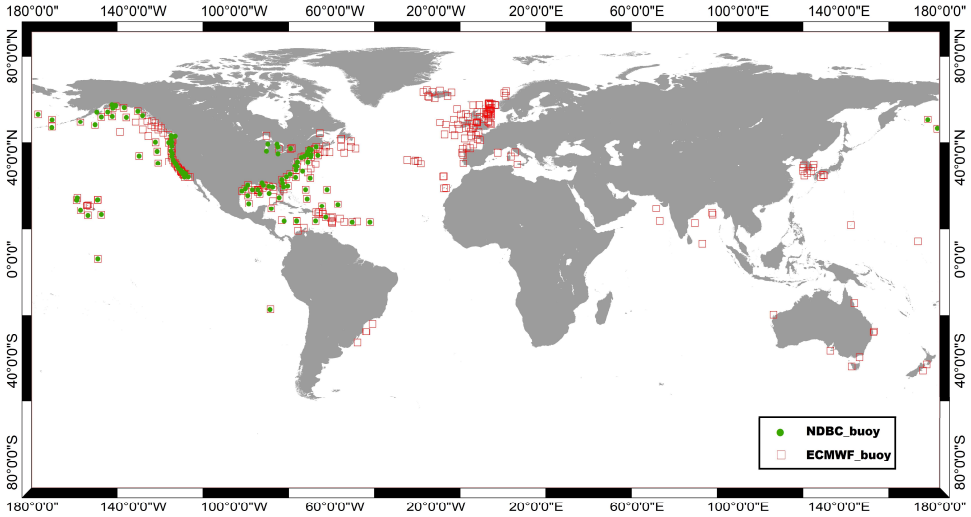


Fig. 2 Locations of the collocated ECMWF buoys (red squares) and NDBC buoys (green dots)

Calibration of the ASAR-derived sea state parameters

Comparison with buoy wave data

To compare the ASAR-derived SWH (denoted ASAR_H_s) with the ECMWF buoy SWH (denoted ECMWF_buoy_H_s), we limited the SWH to the range from 0.5 m to 30.0 m. Eventually, 29,123 data pairs were retained for comparison, and the corresponding scatter diagram is shown in Figure 3(a). Similarly, for a comparison of between the ASAR-derived MWP (denoted ASAR_T_{m02}) and the NDBC buoy MWP (denoted NDBC_T_{m02}), the T_{m02} was limited to the range from 2.0 s to 20.0 s. Eventually, 15,393 data pairs were used for calibration and validation, and the corresponding scatter diagram is shown in Figure 3(b). The colors in the two diagrams indicate the density of data pairs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

The following four statistical parameters were used to evaluate the comparisons of the ASAR-derived (referring to both raw and calibrated) sea state parameters with buoy data or RA data, where x represents the ASAR-derived sea state parameters and y represents either the buoy data or the RA data.

$$\rho(\mathrm{X},\mathrm{Y}) = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

$$Cov(X,Y) = \frac{1}{N}\sum(\mathrm{X}_i - \bar{\mathrm{X}})(\mathrm{Y}_i - \bar{\mathrm{Y}})$$

$$\mathrm{D(X)} = \frac{1}{N}\sum(X_i - \bar{X})^2,\ \mathrm{D(Y)} = \frac{1}{N}\sum(Y_i - \bar{Y})^2$$

$$\mathrm{Bias} = \bar{\mathrm{Y}} - \bar{\mathrm{X}} \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum(Y_i - X_i)^2}{N}}$$

$$\mathrm{SI} = \frac{1}{\bar{Y}}\sqrt{\frac{\sum[(Y_i - \bar{Y}) - (X_i - \bar{X})]^2}{N}}$$

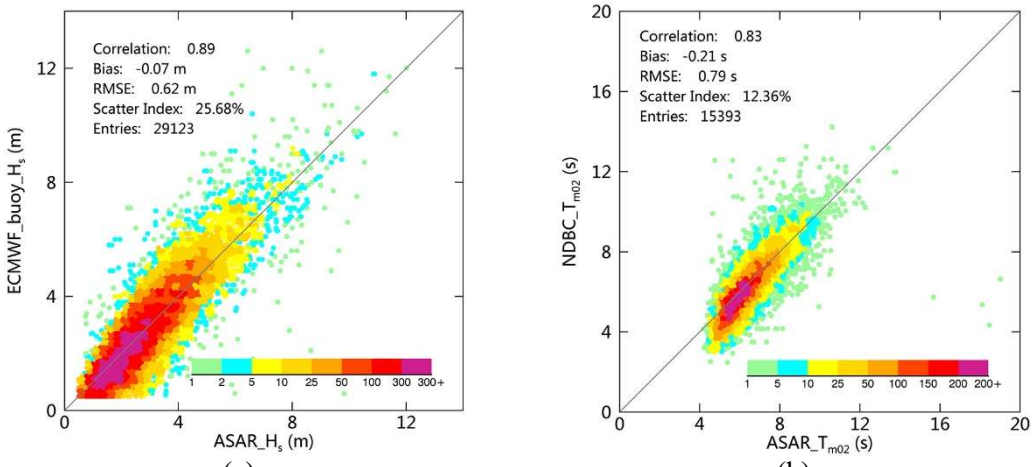


Fig. 3 (a) Comparison between the ASAR-derived SWH and the ECMWF buoy data. (b) Comparison between the ASAR-derived MWP and the NDBC buoy data.

ASAR_H_S is slightly higher than ECMWF_buoy_H_S with a bias of 0.07 m. The RMSE is 0.62 m, which is close to the result (0.70 m) achieved in the preliminary validation based on a two-month dataset[17]. The relatively higher scatter index (S.I.) of 25.68% might be

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

attributable to the hourly resolution of the ECMWF buoy data, whereas the ASAR observations are instantaneous; additionally, the relatively large collocation distance of 100 km may have influenced this result. Furthermore, a comparison of the MWP results suggests that the ASAR retrievals are also slightly higher than the NDBC buoy-measured periods with a bias of 0.21 s. In contrast, the retrieved ASAR_T_{m0} are closely distributed both sides of the 1:1 diagonal line, and therefore, the comparison yields a low S.I. of 12.36%. With respect to the correlation coefficient, both comparisons suggest that the ASAR retrievals display good agreement with the ECMWF and NDBC buoy measurements, having values of 0.89 and 0.83, respectively.

Calibration of the ASAR-derived SWH data

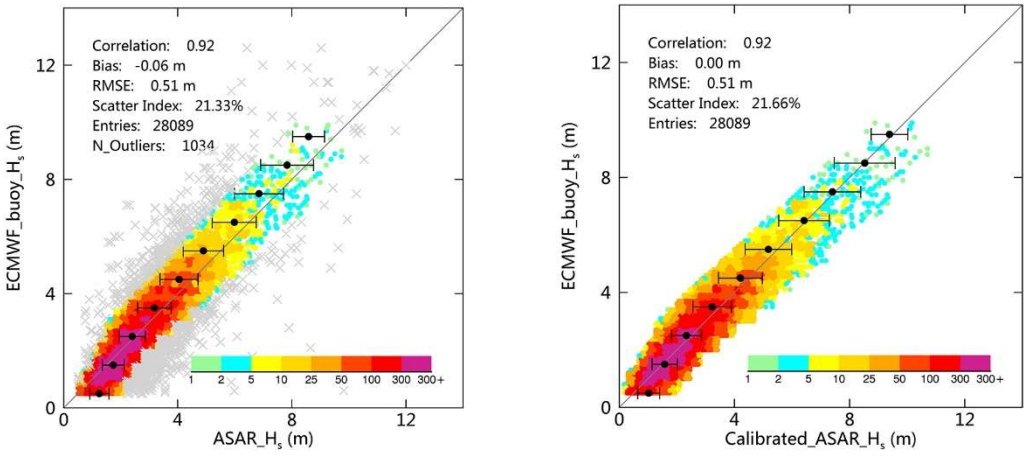
Our goal is to calibrate the ASAR-derived sea state parameters using buoy measurements; however, quite a few collocations are outliers, as illustrated in Figure 3. If these outliers are included in the calibration process, they can introduce uncertainty. Therefore, we used quartiles to exclude some outliers from the calibration process[29]. Quartiles are obtained by dividing the data sorted into ascending order into four equal groups, which can be used to describe the distribution of the data and identify the outliers. The second quartile Q2 is the median of the data. The first quartile Q1 and the third quartile Q3 represent the data between the median and the minimum and maximum, respectively. IQR is the interquartile range. According to Q1, Q2, Q3 and the IQR, the lower and upper bounds can be calculated. The data exceeding the lower and upper fences are regarded as outliers.

$$\begin{aligned} \mathrm{IQR} &= \mathrm{Q3} - \mathrm{Q1} \\ \text{lower bound} &= \mathrm{Q1} - 1.5\mathrm{IQR} \\ \text{Upper bound} &= \mathrm{Q3} + 1.5\mathrm{IQR} \end{aligned} \tag{6}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

By applying these quartiles to exclude some outliers, we further employed robust regression to detect the outliers[30,31] of the collocated data pairs. Robust regression is a linear regression method that is insensitive to outliers. At the start of the regression, all the fitting data have equal weights. By applying least-square minimization, the predicted values and residuals are calculated, where the residuals represent the difference between the predicted values and the observed ones. The data with large residuals are assigned small weights in the subsequent iterations. After a few iterations, the weights of the fitting data are adjusted, and the outliers are verified to have small weights. In this study, the fitting data with weights smaller 0.15 are considered outliers and are excluded from the calibration of the ASAR SWH data.

The cross symbols in Figure 4 represent the outliers detected by the two processing steps described above.



(a) (b)

Fig. 4 (a) Comparison between the ASAR-derived SWH and the ECMWF buoy SWH. (b) Comparison between ASAR-derived SWH and the ECMWF buoy SWH after calibration.

Although the quantile and robust regression methods successfully excluded some data pairs as outliers (as indicted by the improved statistical parameters), the comparison shown in Figure 4(a) suggests that the difference between ASAR_H_S and ECMWF_buoy_H_S is still distinct; specifically, the underestimation of the SWH

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

increases along with sea state varying. In the next step, the buoy measurements were used to calibrate the ASAR retrievals.

The buoy measurements are the best data source of Cal/Val of satellite retrievals. However, these data are not completely unbiased or free of errors[32]. Therefore, we used the reduced major axis (RMA) regression method [31,33], which treats the variables x (ECMWF_buoy_H_s) and y (ASAR_H_s) independently, to calibrate the ASAR retrievals. In the regression, the errors of x and y are both considered by minimizing the triangular area $0.5 * (\Delta x \Delta y)$ between the data points and the regression line, where Δx and Δy are the distances between the actual and predicted values in the x and y directions, respectively. By applying RMA regression to the collocated data pairs, the following linear calibration formula for the ASAR SWH data is obtained:

$$\text{Calibrated\_ASAR\_}H_s(m) = 1.140 * ASAR\_H_s(m) - 0.402 \quad (7)$$

Figure 4(b) shows a comparison between the Calibrated_ASAR_H_s and the buoy measurements ECMWF_buoy_H_s. The calibration process does improve the bias, which decreases from 0.06 m to zero. However, the other three parameters, including the correlation coefficient, RMSE and S.I., do not improve. Although performing the calibration does not improve the overall statistical parameters, it significantly improves the underestimation of the ASAR-retrieved SWH, as revealed by the error bars overlaid on the scatter diagram, while the underestimation trend originally increases with the wave height. Because the collocated data pairs are unequally distributed among different wave heights and much of the data (62.58%) are associated with a low to moderate sea state (SWH< 2.5 m), the overall statistical parameters do not reflect the effect of calibration on the ASAR-retrieved SWH for different sea states. The following

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

table lists the variations in the bias and RMSE with the sea state (the Douglas sea scale is used) before and after applying the RMA calibration to the collocated data pairs.

The bias is significantly reduced by the calibration process, particularly for the slight, and higher than rough sea states. This finding indicates that the linear calibration partially reduces the problem of overestimation for slight to moderate sea states and underestimation for rough and high sea states. The RMSE displays slight fluctuations before and after the calibration process, except for the very high sea state (SWH larger than 9.00 m), for which it is reduced by approximately 33% after the calibration.

Table 1 Variations in the bias (Buoy – ASAR) and RMSE with the sea state before and after RMA calibration

Range (m)	Description	Bias (m)		RMSE (m)	
		Raw	Calibrated	Raw	Calibrated
0.50-1.25	Slight	-0.43	-0.22	0.54	0.44
1.25-2.50	Moderate	-0.16	-0.03	0.40	0.41
2.50-4.00	Rough	0.14	0.13	0.52	0.59
4.00-6.00	Very Rough	0.38	0.18	0.74	0.75
6.00-9.00	High	0.46	-0.04	0.89	0.87
9.00-14.00	Very High	0.74	-0.06	0.92	0.62

Calibration of the ASAR-derived MWP

Following the same calibration method applied to the SWH, the NDBC buoy data are used to calibrate the ASAR-derived MWP. In total, 15,393 data pairs were collected to compare the MWP considering the collocation criteria mentioned above. After elimination of outliers by the quartile and robust regression methods, 14,970 pairs of data remained. The scatter diagram of the comparison is shown in Figure 5(a), where

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

the colors represent the density of data pairs and the cross symbols indicate the detected outliers. Using the RMA regression method, a linear calibration of the MWP is derived:

$$Calibrated\_ASAR\_T_{m02}(s) = 1.268 * ASAR\_T_{m02}(s) - 1.887 \quad (8)$$

The calibrated ASAR MWP results are plotted against the NDBC buoy data in Figure 5(b). Comparing Figure 5 (a) with (b), the calibration improves both the bias and the RMSE, which decrease from -0.19 s to zero and from 0.67 s to 0.65 s, respectively. However, the correlation coefficient and S.I. do not improve. The raw data suggest that the ASAR-derived results overestimate the MWP below 7 s but underestimate it above 8 s. The calibration makes the data pairs almost symmetrically distributed about the 1:1 diagonal line and partially corrects the trend result.

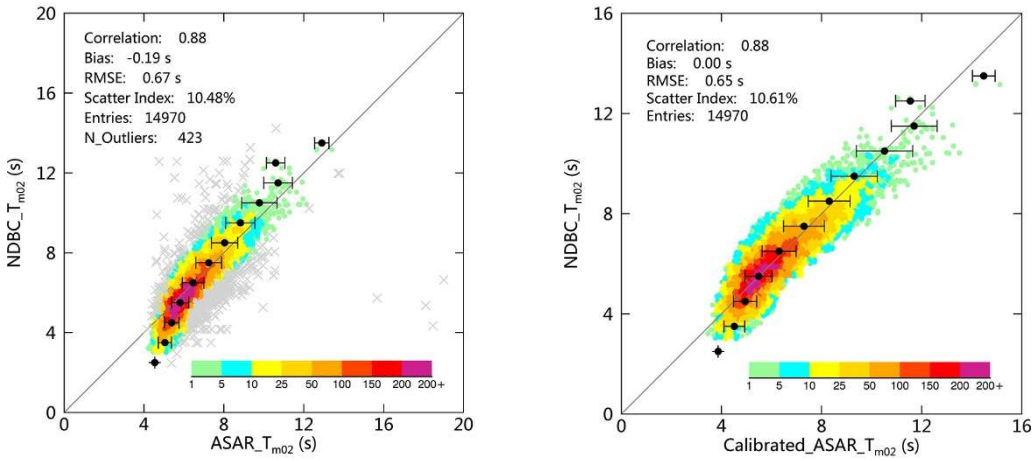


Fig. 5 (a) Scatter diagram of the comparison between the ASAR-derived MWP and the NDBC buoy MWP. (b) The same as (a) but for calibrated ASAR-derived MWP.

RA wave data

The GlobWave project also collected wind and wave data for the Geodetic Satellite (GEOSAT), GEOSAT Follow-on (GFO), ERS-1, ERS-2, TOPEX/POSEIDON, JASON-1, JASON-2 and CryoSAT-2 RA missions, with a time span from 1985

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

onwards. The Jason-1 mission provided wave data from December 2001 until July 2013, which covers the lifetime of the ASAR instrument. GlobWave reprocessed the original JASON-1 measurements and provided quality control flags and calibrated SWH measurements. In this study, we used calibrated Ku-band SWH measurements of JASON-1 to perform a cross-validation with the calibrated ASAR-derived SWH. A quality flag named '*swh_quality*' provided in the GlobWave RA products is used to filter the JASON-1 SWH data with high quality for validation. This flag has three values, namely, 0, 1, and 2, representing a '*good_measurement*', '*acceptable_for_some_applications*' and a '*bad_measurement*', respectively. Only the data flagged as '*good_measurement*' are used for validation. The same collocation criteria employed in the collocation of buoy data were utilized between the ASAR data and the JASON-1 data.

Data records

The ASAR WM data global wave product is stored in NetCDF-3 format and follows the Climate and Forecast Metadata CF-1.7 convention[34]. The naming convention of the ASAR sea state product files are as follows:

Satid_Sensor_Type_StartDate_StartTime_EndDate_EndTime_Cycle_Orbit.NC,

where

a. Satid: mission name
b. Sensor: sensor name
c. Type: type of product
d. StartDate: Date of the first record
e. StartTime: Time of the first record
f. EndDate: Date of the last record

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

g. EndTime: Time of the last record

h. Cycle: cycle number of the satellite

i. Orbit: relative orbit number of the satellite

The records contained in the product correspond to the imagettes of the ASAR WVI Level 1B product. Each record consists of 14 variables, which are listed in the following table.

Table 2 List of variables and their descriptions in the ASAR WM sea state NetCDF product

No.	Variables	Description
1	Time	Acquisition time of the ASAR imagettes. Seconds since 2000-01-01 00:00:00 UTC
2	Latitude	Latitude of ASAR imagette center
3	Longitude	Longitude of ASAR imagette center
4	Heading	Flight direction of the satellite (clockwise relative to north)
5	Inci_angle	Local incidence angle of ASAR imagette center
6	Homogeneity	Homogeneity of ASAR imagettes
7	SWH	Retrieved SWH of ASAR imagettes
8	MWP	Retrieved zero-upcrossing wave period of ASAR imagettes
9	SWH_Cali	Calibrated SWH
10	MWP_Cali	Calibrated zero-upcrossing MWP
11	Rejection_Flag	The records flagged 0B are acceptable records 1B for a 'bad_record' 2B for 'land' 3B for 'inhomogeneous ASAR imagettes' 4B for 'ASAR imagettes in HH polarization' and 5B for 'ASAR imagettes with an incidence angle not equal to 23°' 6B for 'ASAR imagettes in the polar regions, i.e. beyond 70°N or 65°S'

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

12	Land_Flag	0B for ocean area 1B for land area
13	Normalized_variance	Normalized variance of an SAR image
14	QC_Flag	0B for a good record 1B for a suspect record 2B for a bad record 3B for an unprocessed record

'*SWH*' and '*MWP*' are the retrieved results using the CWAVE_ENV model. By applying the calibration formulas given in equations 7 and 8, the calibrated ASAR-derived SWH and MWP are obtained and stored as the variables '*SWH_Cali*' and '*MWP_Cali*'.

The '*Rejection_Flag*' flags the data with values of 0, 1, 2, 3, 4, 5 or 6, which represent an acceptable record, a bad record, a record containing land (discrimination is based on '*Land_flag*'), an inhomogeneous (homogeneity factor $\geqslant$ 1.05) ASAR imagette, an imagette acquired in HH polarization, and an imagette with an incidence angle not equal to 23°, an imagetted acquired in the polar regions, respectively. The '*Land_flag*' is transformed from the ASAR WM Level 1B data. The '*Normalized_variance*'[25] variable is the normalized variance of the ASAR WM intensity data and is calculated according to equation 9.

$$Noramalized\_Variance = \frac{I_{var}}{I_{mean} * I_{mean}}$$

$$I_{mean} = \frac{\sum_{i,j=1}^{M,N} I_{i,j}}{M * N} \quad (9)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

$$I_{var} = \frac{\sum_{i,j=1}^{M,N}\left(I_{i,j} - I_{mean}\right)^2}{M * N}$$

where I_{var} and I_{mean} represent the variance and mean of the image, respectively, and M and N refer to the width and height of the image, respectively.

The '*QC_Flag*' variable has four values that describe the quality of a record. We considered a few factors during the quality control process, including the reasonable range of variables, the normalized variance and the '*Rejection_flag*'. Based on the factors, the records were divided into 4 levels.

(1) Good record (*QC_Flag* = 0B), which satisfies the following criteria:

a. 0.5 m $\leqslant$ *SWH* (*SWH_Cali*) < 30 m and 0 s < *MWP* (*MWP_Cali*) < 20 s

b. $\overline{\sigma_0}$ -NESZ > 3 dB

c. *Rejection_Flag = 0B*

where $\overline{\sigma_0}$ is the mean normalized radar cross section of the ASAR imagettes and NESZ is the noise equivalent sigma zero, i.e., the noise floor of the ASAR WM data.

(2) Suspected record (*QC_Flag* = 1B), which satisfies the following criterion:

a. SWH > 30 m or MWP > 20 s

(3) Bad record (*QC_Flag* = 2B), which satisfies one of the following conditions:

a. *SWH (SWH_Cali)* < 0 m or *MWP* (*MWP_Cali*) < 0 s

b. $\overline{\sigma_0}$ -NESZ $\leqslant$ 3 dB

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

(4) Any record with the variable '*Rejection_Flag*' not equal to 0B is classified as an unprocessed record (*QC_Flag* = 3B). These ASAR WM data, which are excluded from further processing, and the values of all the corresponding variables are set to '*Fillvalue*'.

Technical validation

Comparison with RA wave data

Following the collocation criteria, 46,642 data pairs of JASON-1 and ASAR WM data were obtained. However, a large number of SWH measurements from JASON-1 of the GlobWave product in 2012 were abnormal, ranging from -40 m to 40 m and exhibiting a discontinuous spatial distribution. These data were discarded from the validation dataset. In addition, we set a valid range of SWH from 0.5 m to 30 m for validation. Finally, 23,192 pairs of JASON-1 and ASAR WM data were obtained. Using the quartile method described above to exclude outliers, 22,862 pairs of data were collected for validation. Figure 6(a) shows the comparison between the ASAR-derived SWH and JASON-1 SWH (denoted JASON1_H_s). The robust regression method was not applied to exclude outliers because we consider both datasets to comprise independent measurements. The calibrated ASAR SWH (applying equation 7) is also compared with the JASON-1 calibrated SWH[35] (Calibrated_JASON1_H_s), as shown in Figure 6(b).

As shown in Figure 6(a), the ASAR SWH displays good consistency with the JASON-1 SWH, and the bias and RMSE are 0.04 m and 0.48 m, respectively; additionally, the correlation coefficient and S.I. are 0.93 and 16.84%, respectively. Although the ASAR SWH is generally slightly lower than the JASON-1 SWH, it is higher for a relatively low sea state (SWH < 2.5 m). In Figure 6(b), the calibrated ASAR SWH also displays good agreement with the calibrated JASON-1 SWH, with bias, RMSE, correlation coefficient and S.I. values of 0.18 m, 0.53 m, 0.93 and 16.64%, respectively. The Q-Q plots shown in Figure 6 (c) and (d) suggest that the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

underestimation of ASAR-derived SWH is significantly improved after the calibration process, particularly for SWH above 6 m.

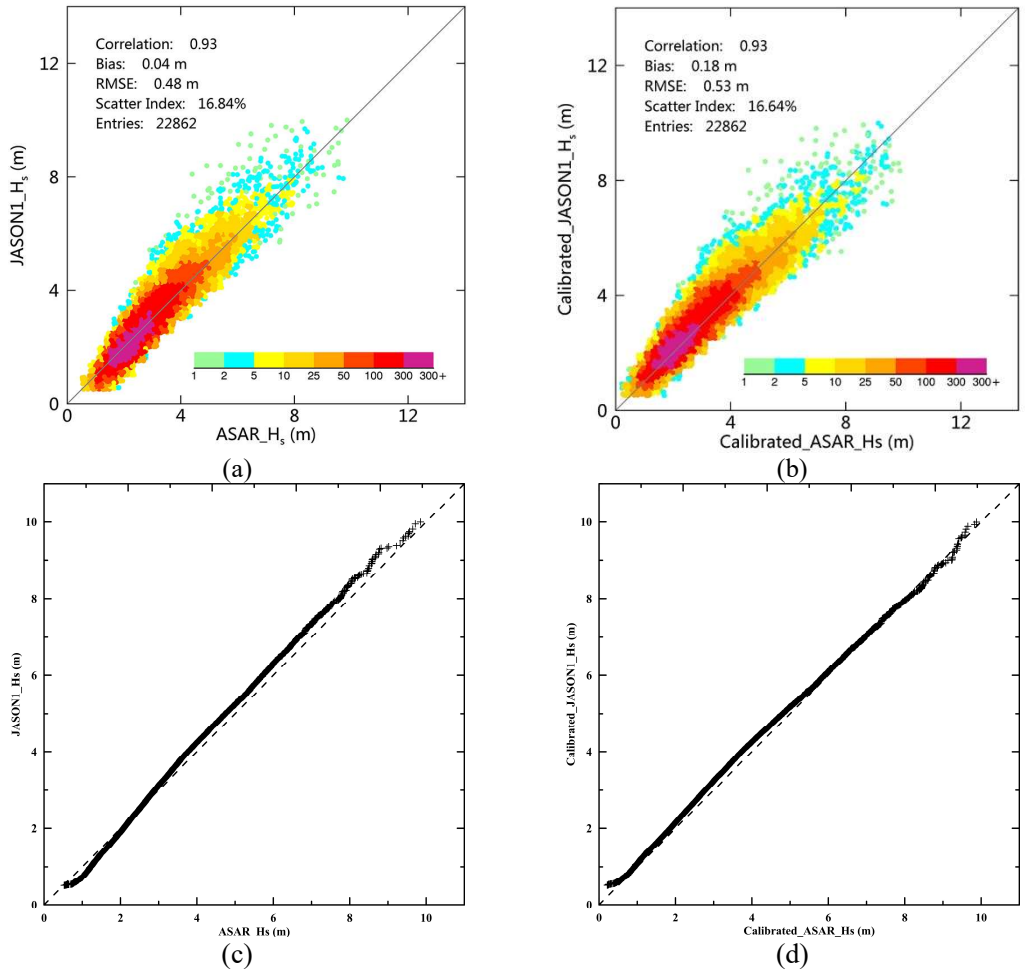


Fig. 6 (a) and (c) Comparison between the ASAR-derived SWH and the RA SWH and the corresponding Q-Q plot, respectively. (b) Comparison between the calibrated ASAR-derived SWH and the calibrated RA SWH and the corresponding Q-Q plot, respectively.

A major limitation of these overall comparisons in evaluating the retrieval of sea state parameters is that the data pairs are unevenly distributed among different sea states. As the sea state increases in severity, the number of valid data pairs decreases. Therefore, a stepwise comparison was conducted to assess the performance of the ASAR SWH data quality for different sea states. Figure 7(a) shows the uncalibrated and calibrated ASAR SWH compared with the JASON-1 SWH at a 1-m interval. Figure 7 (b) is the same as (a) but compares the ASAR SWH with the calibrated JASON-1 SWH.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 January 2020 doi:10.20944/preprints202001.0200.v1

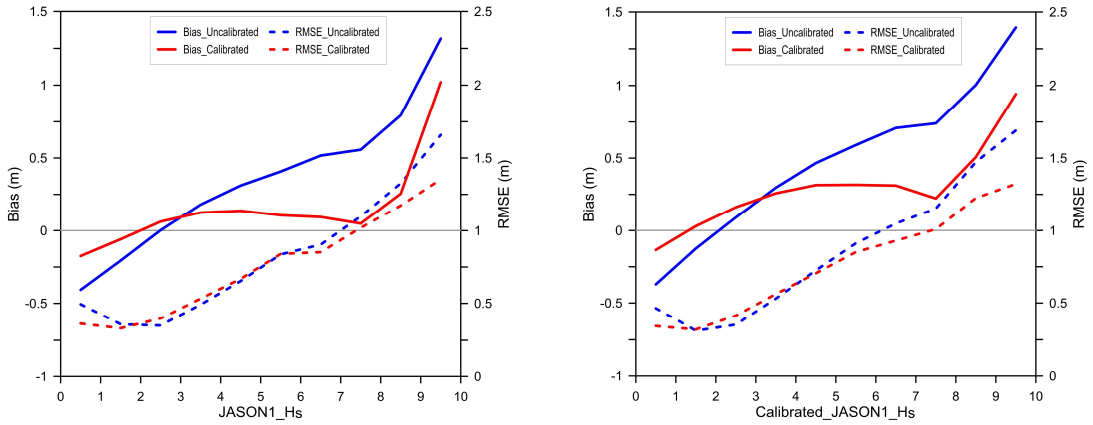


Fig. 7 Variations in the bias and RMSE of the ASAR-derived SWH versus the RA SWH (a) and calibrated SWH (b).

Because the changes in the bias and RMSE illuminated in Figure 7(a) and (b) show similar trends, we use Figure 7(a) as an example. Figure 7(a) shows the changes in the bias and RMSE of the uncalibrated and calibrated ASAR SWH versus the JASON-1 SWH, where the solid and dashed lines represent the bias and RMSE, respectively and the red and blue lines refer to the calibrated and uncalibrated ASAR SWH, respectively. The bias of the uncalibrated ASAR SWH increases with the sea state and changes from negative to positive when the SWH is approximately 2 m. The calibration process significantly reduces the bias to less than 0.15 to 0.2 m from low to high sea states (at approximately 8 m), and importantly, the bias becomes less dependent on the sea state increasing. For a very high sea state (SWH>9 m), the bias accounts for approximately 10% of the total SWH; additionally, the RMSE of the calibrated ASAR SWH varies from 0.25 m to 1.20 m and is particularly reduced for sea states higher than very rough (above approximately 5 m).

A comparison of different measurements or retrievals for a high sea state is always difficult. On the one hand, collocations for high sea states are rare, leading to uncertainties. On the other hand, high sea states can exhibit significant spatial and

temporal variations, while collocations of ASAR WM and JASON-1 data are in within 100 km distance and 0.5 hours.

Code availability

The IDL code for reading the ocean wave parameter products is provided. The code for producing the products is available upon request through scientific cooperation.

Data Citation

1. Li Xiao-Ming, Huang Bingqing (2020). A global sea state dataset from spaceborne synthetic aperture radar wave mode data. SEANOE. https://doi.org/10.17882/71337

Acknowledgments

The first author would like to thank Dr. Susanne Lehner and Dr. Johannes Schulz-Stellenfleth, my supervisor and former colleague for their supports to develop the CWAVE_ENV parametric model for the ASAR wave mode data. We particularly thank the ESA for agreeing to deliver us the entire ASAR wave mode dataset under the framework of "Dragon 4 program". Mr. Jean-Francois Piollo from Ifremer/Cersat gave us the cordial help of delivering these wave mode data. Dr. Alexis Mouche, also from Ifremer/Cersat, encouraged us to participate the ESA sea state CCI project. This is one of our motivations to reprocess the entire ASAR WM dataset for times to standard ocean wave products. The study is partially supported by the National Key Research and Development Project (2018YFC1407100).

Author Contributions

X.M. initiated, designed and implemented the project and performed the bulk of the work creating the dataset, conducting the evaluations and writing the paper. B.Q. worked on the software to produce the products and performed the calibration and validation tasks. She also contributed to drafting the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Ribal, A. & Young, I. R. 33 years of globally calibrated wave height and wind speed data based on altimeter observations. *Scientific Data*. **6,** 1-15 (2019)
2. Young, I. R., Zieger, S., & Babanin, A. V. Global trends in wind speed and wave height. *Science*. **332,** 451-455 (2011),
3. Brown, W. M. & Procello, C. J. *An introduction to synthetic aperture radar*, IEEE spectrum. **6,** 57-62 (1969).
4. Born, G. H., Dunne, J.A. & Lame, D. B. Seasat mission overview, *Science*. **204,** 1405 – 1406 (1979)
5. McLeish, W., Ross, D., Shuchman, R. A., Teleki, P. G., Hsiao, S. V., Shemdin, O. H., & Brown, W. E. Synthetic Aperture Radar imaging of ocean waves: Comparison with wave measurements. *J. Geophys. Res.* **85,** 5003– 5011 (1980).
6. Alpers, W. & Rufenach, C. L. The effect of orbital motions of Synthetic Aperture Radar imagery of ocean waves. *IEEE transactions on Antennas and Propagation*. **27,** 685–690 (1979).
7. Hasselmann, K., Raney, R. K., Plant, W. J., Alpers, W., Shuchman, R. A., Lyzenga, D. R., Rufenach, C. L., & Tucker, M. J., Theory of Synthetic Aperture Radar ocean imaging: A MARSEN view. *J. Geophys. Res.* **90,** 4659– 4686 (1985).
8. Hasselmann, K. & Hasselmann, S. On the nonlinear mapping of an ocean wave spectrum into a synthetic aperture radar image spectrum. *J. Geophys. Res.* **96,** 10713-10729 (1991).
9. WAMDI GROUP. The WAM model a third generation ocean wave prediction model. *J. Phys. Oceanogry.* **18,** 1775-1810 (1984).
10. Mastenbroek, C. & de Valk, C. A semi-parametric algorithm to retrieve ocean wave spectra from synthetic aperture radar. *J. Geophys. Res.* **105,** 3497-3516 (1998).
11. Hasselmann, S., Brüning, C., Hasselmann, K., & Heimbach, P. An improved algorithm for the retrieval of ocean wave spectra from synthetic aperture radar image spectra. *J. Geophys. Res.* **101,** 16615– 16629 (1996).
12. Kerbaol, V., Chapron, B. & Vachon, P. W. Analysis of ERS-1/2 synthetic aperture radar wave mode imagettes. *J. Geophys. Res.* **103,** 7833–7846 (1998).
13. Lehner, S., Schulz-Stellenfleth, J., Schättler, J. B., Breit, H. & Horstmann, J. Wind and wave measurements using complex ERS-2 wave mode data. *IEEE Trans. Geosci., and Rem. Sens.* **38,** 2246-2257 (2000).
14. Schulz-Stellenfleth, J., König, Th. & Lehner, S. An empirical approach for the retrieval of integral ocean wave parameters from synthetic aperture radar data. *J. Geophys. Res.* **112,** (2007).
15. Hasselmann, K. et al. The ERS SAR wave mode: a breakthrough in global ocean wave observations. *ESA special publication.* **SP-1326,** 167-197, (2012).
16. Li, X. M., König, T., Schulz-Stellenfleth, J. & Lehner, S. Validation and intercomparison of ocean wave spectra retrieval scheme using ASAR wave mode data. *International Journal of Remote Sensing*. **31,** 4969-4993 (2010).
17. Li, X. M., Lehner, S. & Bruns, T. Ocean wave integral parameter measurements using Envisat ASAR wave mode Data. *IEEE Trans. Geosci. Remote Sens.* **49,** 155-174 (2011).
18. Stopa, J. E. & A. Mouche. Significant wave heights from Sentinel-1 SAR: Validation and applications. *J. Geophys. Res.* **122,** 1827-1848 (2017).
19. Li, X. M., Zhang, T. Y., Huang, B. Q., & Jia, T. Capabilities of Chinese Gaofen-3 synthetic aperture radar in selected topics for coastal and ocean observations. *Remote Sens.* **10,** 1-22 (2018).
20. Ardhuin, F., Chapron, B. & Collard, F. Observation of swell dissipation across oceans. *Geophys. Res. Lett.* **36,** L06607, (2009).
21. Collard, F., Ardhuin, F. & Chapron, B. Monitoring and analysis of ocean swell fields from

space: New methods for routine observations. *J. Geophys. Res.* **114,** C07023 (2009).
22. Li, X. M. A new insight from space into swell propagation and crossing in the global oceans. *Geophys. Res. Lett.* **43,** 5202-5209 (2016).
23. Chapron, B., Johnson, H. & Garello, R. Wave and wind retrieval from SAR images of the ocean, *Ann. Telecommun.* **56,** 682−699 (2001).
24. Engen, G. & Johnson, H. SAR-ocean wave inversion using image cross spectra. *IEEE Trans. Geosci. Remote Sens.* **33,** 329–360 (2000).
25. Brooker, G. UWA processing algorithm specification. version 2.0, Tech. Rep., ESA, ESTEC/NWP, Noordwijk, The Netherland, 1995.
26. European Space Agency, ENVISAT ASAR Product Handbook, Issue 2.2, 2007.
27. Stoffelen, A. & Anderson, D. Scatterometer data interpretation: Estimation and validation of the transfer function CMOD4. *J. Geophys. Res.* **102,** 5767– 5780 (1997).
28. Schulz-Stellenfleth, J. & Lehner, S. Measurement of 2-D sea surface elevation fields using complex Synthetic Aperture Radar data. *IEEE Trans. Geosci., & Rem. Sens.* **42,** 1149-1160 (2004).
29. Tukey, J. *Exploratory Data Analysis*. (Addison-Wesley. 1977)
30. Rousseeuw, P. & Leroy, A. *Robust Regression and Outlier Detection 3rd edn* (John Wiley & Sons, 1996)
31. Zieger, S., Vinoth, J., & Young, I. R. Joint calibration of multiplatform altimeter measurements of wind speed and wave height over the past 20 years. *Journal of Atmospheric and Oceanic Technology*. **26,** 2549-2564 (2009).
32. Stoffelen, A. Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *J. Geophys.Res.* **103,** 7755–7766 (1998).
33. Trauth, M. MATLAB Recipes for Earth Sciences. 4th Edn (Springer, 2015)
34. Eaton, B., Gregory, J. & Bod Drach etc. NetCDF Climate and Forecast (CF) Metadata Convebtions Version 1.7.