Acta Oceanol. Sin., 2019, Vol. 38, No. 9, P. 120–127
https://doi.org/10.1007/s13131-019-1480-2
http://www.hyxb.org.cn
E-mail: hyxbe@263.net

A newly developed ocean significant wave height retrieval method from Envisat ASAR wave mode imagery

Chenqing Fan[1], Xiaochen Wang[2, 3*], Xudong Zhang[1, 4], Dong Gao[1]

[1] First Institute of Oceanography, Ministry of Natural Resources, Qingdao 266061, China

[2] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China

[3] Laboratory of Target Microwave Properties, Deqing Academy of Satellite Applications, Deqing 313200, China

[4] College of information and engineering, Ocean University of China, Qingdao 266071, China

Received 19 January 2018; accepted 1 May 2018

Abstract

The main objective of this paper is to propose a newly developed ocean Significant Wave Height (SWH) retrieval method from Envisat Advanced Synthetic Aperture Radar (ASAR) imagery. A series of wave mode imagery from January, April and May of 2011 are collocated with ERA-Interim reanalysis SWH data. Based on the matched datasets, a simplified empirical relationship between 22 types of SAR imagery parameters and SWH products is developed with the Genetic Algorithms Partial Least-Squares (GA-PLS) model. Two major features of the backscattering coefficient σ_0 and the frequency parameter S_{10} are chosen as the optimal training feature subset of SWH retrieval by using cross validation. In addition, we also present a comparison of the retrieval results of the simplified empirical relationship with the collocated ERA-Interim data. The results show that the assessment index of the correlation coefficient, the bias, the root-mean-square error of cross validation (RMSECV) and the scattering index (SI) are 0.78, 0.07 m, 0.76 m and 0.5, respectively. In addition, the comparison of the retrieved SWH data between our simplifying model and the Jason-2 radar altimeter data is proposed in our study. Moreover, we also make a comparison of the retrieval of SWH data between our developed model and the well-known CWAVE_ENV model. The results show that satisfying retrieval results are acquired in the low-moderate sea state, but major bias appears in the high sea state, especially for SWH>5 m.

Key words: significant wave height, Envisat ASAR, GA-PLS, optimal feature subset

Citation: Fan Chenqing, Wang Xiaochen, Zhang Xudong, Gao Dong. 2019. A newly developed ocean significant wave height retrieval method from Envisat ASAR wave mode imagery. Acta Oceanologica Sinica, 38(9): 120–127, doi: 10.1007/s13131-019-1480-2

1 Introduction

The Significant Wave Height (SWH) is a noteworthy ocean surface parameter and is shown to be of great importance to ocean dynamics monitoring, disaster forecasting and off-shore engineering. With the advantage of high spatial resolution, a relatively large spatial coverage and the capability of imaging under almost all weather, day-and-night conditions, the Synthetic Aperture Radar (SAR) has great potential in ocean surface monitoring, especially for SWH (Gonzalez et al., 1979; Longuet-Higgins et al., 1963).

Generally, there are two main approaches for the present measurement of SWH from SAR data. One approach is to build the non-linear analytic model from the SAR image spectrum to the ocean wave spectrum, and after that, the SWH can be derived from the ocean wave spectrum by spectrum integration (Long, 1980). For example, Hasselmann and Hasselmann (1991) developed the Max Planck Institute (MPI) algorithm on the basis of the three modulations of tilt, hydrodynamics and velocity bunching. However, only the precise initial guess spectrum from the wave model as an input can efficiently retrieve ocean information, which has limited its application, especially in off-shore applications. Considering the difficulty in acquiring the precise initial guess spectrum, He et al. (2004) developed a parameterization algorithm to eliminate the influence of initial guess spectra in which the parameters of the ocean spectrum are determined from the SAR image spectrum directly, and the 180° ambiguity of ocean propagation is dismissed by the two sub-images with different incidence angles. In fact, the two SAR sub-image spectra are not actually consistent, and the retrieval results of mixed-waves are not satisfactory. In addition, Mastenbroek and de Valk (2000) proposed the Semi-Parametric Retrieval Algorithm (SPRA) by separating the wind-wave and swell-wave spectra with a synchronous scatterometer and SAR. However, the SPRA is not rigorously based on the non-linear imaging relationship, which results in discontinuity between the measured shortest wave and the parameterization wind-wave spectrum. The 180° ambiguity of ocean propagation is also not well settled. Enegn and Johnson (Engen and Johnsen, 1995) took two looks at SAR Single Look Complex (SLC) imagery in which the cross spectrum can be derived to remove the 180° ambiguity of the ocean wave propagation direction, which is named the cross-spectrum algorithm. Although the cross-spectrum algorithm can dismiss the 180° ambiguity of ocean wave propagation, the retrieved ocean wave spectrum presents a major bias due to the absence of synchronous

Foundation item: The National Science Foundation for Young Scientists of China under contract No. 61501130; the National Key Research and Development Program of China under contract Nos 2016YFB0502504 and 2016YFB0502500; the National Natural Science Foundation of China under contract Nos 41431174, 61471358 and 41401427.

*Corresponding author, E-mail: wangxc@aircas.ac.cn

wind field information, in which the calculated wind speed from the CMOD model uses a supposed wind direction of 45°.

Recently, Wang et al. (2012) proposed a semi-empirical algorithm that estimates the wave height from Envisat ASAR wave mode imagery without any prior knowledge. Shao et al. (2016) developed a semi-empirical function for significant wave height (SWH) and mean wave period retrieval from C-band VV-polarization Sentinel-1 SAR. Collard et al. (2005) adapted the wave spectrum inversion scheme for shallow water from the European Space Agency (ESA)'s operational processing techniques used for Level 2 ocean wave products. The retrieval results show that there is good overall agreement between the *in situ* and SAR observations under the low to moderate observed wind speed conditions (Wang et al., 2014, 2016). Although major progress has been made in ocean wave spectrum retrieval schemes, the operational retrieval remains a challenge for SAR (Bruck and Lehner, 2015; Grieco et al., 2016). The difficulty in the acquisition of high-precision first guess spectra, the uncertainty of the non-linear relationship between the SAR spectrum and the ocean wave spectrum, and the extra input of wind field information have all limited the applications of the above approaches (Pleskachevsky et al., 2016; Shao et al., 2017).

In addition, it has been verified that the potential of SWH measurement directly from SAR imagery parameters rather than retrieved ocean wave spectrum is feasible. Schhulz-Stellenfleth et al. (2007) proposed an algorithm from the ERS-2 SAR wave mode data that directly builds the empirical relationship between SAR imagery parameters and the ocean wave SWH without any extra input. A good overall agreement is found under low and medium sea states, but there is major bias under the high sea state. Afterwards, Li et al. (2011) improved Schhulz's algorithm from the Envisat ASAR wave mode data. Stopa and Mouche (2017) also developed empirical algorithms for wave mode images measured from the SAR aboard Sentinel-1 A, which are called CWAVE _ ENV and CWAVE_S1A. Unlike Schhulz, Li et al. (2011) built the CWAVE _ ENV algorithm of two-order multiple regression models with 22 types of imagery parameters. Stopa followed the approach of Schulz-Stellenfleth but used neural networks to relate the nonlinear relationships between the seven types of SAR image parameters and output geophysical wave parameters.

In view of algorithm efficiency, the retrieval precision of SWH depends on the number of parameters and the training model. Therefore, it is significant to assess the impact of each parameter on the modeling process and select the weighted feature. Experiments show that properly modifying the Genetic Algorithm (GA) can be an effective feature selection algorithm (Leardi and González, 1998; Song et al., 2012). Compared with the traditional feature selection methods, the independent variables selected by the genetic algorithm can produce better modeling results while using fewer independent variables. Adding the genetic algorithm into the partial least squares algorithm (Geladi and Kowalski, 1986) can simplify the regression model, and can greatly improve the modeling precision of the partial least squares algorithm (Li et al., 2007).

In this paper, we follow the retrieval scheme of CWAVE_ENV and simplify and rebuild the empirical relationship with the GA-PLS model. The ASAR wave mode data of January, April and May of 2011 are collected, and we also present a validation by using the collocated Jason-2 radar altimeter, the ERA-Interim SWH products and the CWAVE_ENV algorithm. In Section 2, we briefly introduce the dataset and the GA-PLS model used in this study. In Section 3, we propose the empirical model based on the GA-PLS model. Moreover, we also make a comparison of the SWH retrieval precision between other match-ups. The conclusions are given in Section 4.

2 Data sets and methodology

2.1 *Data sets*

The data sets used in this paper are the Envisat ASAR wave mode imagery, the Jason-2 Altimeter and the ECMWF reanalysis SWH data.

2.1.1 *Envisat ASAR wave mode*

Approximately 59 974 scenes of Single Look Complex (SLC) imagery are collected from January, April and May of 2011. The former are used to train the construction of the empirical relationship between imagery parameters and SWH, and the latter are used for independent verification. It is noteworthy that the homogeneity check is necessary to exclude the non-wave SAR images before data processing. The parameters of the Envisat ASAR wave mode SLC details are listed in Table 1.

Table 1. Parameters of Envisat ASAR wave mode SLC imagery

Parameter	Envisat ASAR
Operation mode	wave mode
Radar center frequency	5.3 GHz
Polarization	VV
Incidence angle	23°
Resolution	10 m
Swath width	5 km
Imaging interval	100 km

2.1.2 *Jason-2 radar altimeter*

The Ocean Surface Topography Mission (OSTM)/Jason-2 satellite was launched by NASA (USA) and the French Space Agency on June 20, 2008. The Jason-2 radar altimeter also works on a dual-frequency at 13.575 GHz (Ku band) and 5.3 GHz (C band) (Monaldo, 1988). All the along-track SWH data used in this paper are downloaded from AVISO/CNES (http://www.aviso.altimetry.fr/en/data.html). Approximately 2 057 scenes of Jason-2 radar altimeter data are collocated with the ASAR wave mode data with a spatial window of 100 km and a temporal window of 2 h. The location of the matched Jason-2 radar altimeter with the ASAR wave mode data is shown in Fig. 1.

2.1.3 *ERA-Interim*

The ERA-Interim reanalysis SWH data are provided by European Centre for Medium-Range Weather Forecasts (ECMWF) (Wan et al., 2015). The temporal resolution is 6 h, and spatial resolution is 0.125°×0.125°. Approximately 59 974 matched datasets are collected from January, April and May of 2011 and downloaded from the ECMWF (http://apps.ecmwf.int/datasets/). The SWH distribution histogram collocated with the ASAR of January of 2011 and April and May of 2011 are shown in Fig. 2. It is noted that the datasets of January of 2011 are used as training examples, and April and May of 2011 are used as verification examples.

2.2 *Methodology: GA-PLS model*

The GA-PLS model combines the genetic algorithm (GA) and the partial least squares (PLS) algorithm, which can be used to build an empirical relationship between the dependent variable and multi-independent variables (Kawamura et al., 2010). By being combined with the traditional PLS algorithm, the GA-PLS

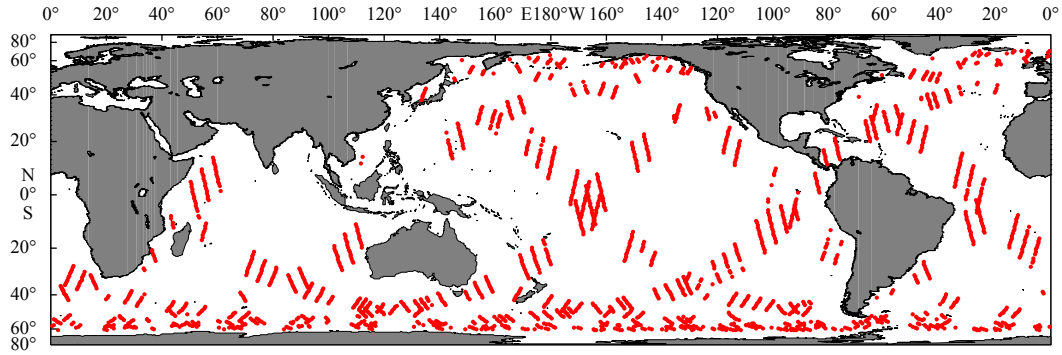

Fig. 1. Location of matched Jason-2 radar altimeter with ASAR wave mode data.

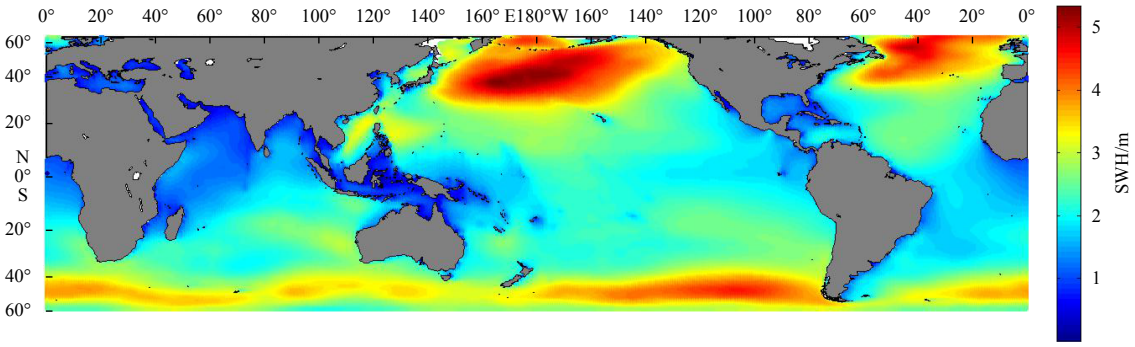

Fig. 2. Global SWH distribution of the ERA-Interim in January, 2011.

model can retain the simplicity of the regression model by adding the feature selection method for optimal feature subset selection. The GA-PLS algorithm can also greatly improve the modeling precision of the partial least squares algorithm with the optimal feature subset. In the GA-PLS modeling process, the independent variables that are closely related to the dependent variable are first removed by the genetic algorithms, and then the concrete mathematical model of a multiple regression can be established by the partial least squares algorithm (Li et al., 2016). It is noted that $\boldsymbol{X}$ represents the multi-independent variables matrix and $\boldsymbol{Y}$ represents the dependent variable matrix. Before PLS, first the principal component analysis (PCA) is applied to $\boldsymbol{X}$ and $\boldsymbol{Y}$.

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}' + \boldsymbol{E}, \tag{1}$$

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Q}' + \boldsymbol{F}, \tag{2}$$

where $\boldsymbol{T}$ and $\boldsymbol{U}$ are the principal component matrices, $\boldsymbol{P}$ and $\boldsymbol{Q}$ are the loading matrices, and $\boldsymbol{E}$ and $\boldsymbol{F}$ are the residual matrices.

The regression equation between the principal component $\boldsymbol{T}$ and $\boldsymbol{U}$ can be built as follows:

$$\boldsymbol{U} = \boldsymbol{B}\boldsymbol{T}, \tag{3}$$

where $\boldsymbol{B}$ is the regression coefficient matrix. The cross-validation method that leaves out one sample at a time was used in this research to select the optimal number of components for the PLS model.

Finally, the multi-regression model is given as follows:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{c}, \tag{4}$$

where $\boldsymbol{b}$ is the final regression coefficient vector and $\boldsymbol{c}$ is the residual vector. The root mean square error of cross validation (RMSECV) is also used to evaluate the precision of the GA-PLS model.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i' - y_i\right)^2}{N}}. \tag{5}$$

3 Results

In this section, the ASAR wave mode imagery matched with the ERA-interim reanalysis SWH data are used for training examples, and another 42 774 scenes are used for validation. Moreover, we also present a comparison of the GA-PLS retrieval results with the Jason-2 radar altimeter and the well-known CWAVE algorithm.

3.1 *An Emperical model for SWH retrieval*

After the time and spatial matching of 17 200 datasets, 22 types of imagery parameters (Table 2) are extracted and applied to build the empirical model as the input training parameters, including the backscattering coefficient σ_0, the imagery variance CVAR and 10 types of frequency domain parameters S. The flow chart of SWH retrieval using the GA-PLS model is shown in Fig. 3.

Following the flow chart (Fig. 3), the GA-PLS model was applied to the data set of matched training samples with the aim to reduce the number of field variables and increase the prediction power of the model. Figure 4 shows the smoothed frequency of the selection for each parameter. It is noticed that the smoothed frequency of selection is defined as a weighted average between the frequency of selection of the variables in the starting run and

Table 2. Imagery parameters used for model training

Parameter	Defination
Backscattering coefficient	$\sigma_0 = 10\lg(DN_{ij})$
Imagery variance	$CVAR = \left(\mathrm{std}\left(\frac{A-\langle A\rangle}{\langle A\rangle}\right)\right)^2$
Frequency domain parameter	$S_i = \int_A \bar{P}(k_x, k_y)\,\overline{h_i}(k_x, k_y)\,\mathrm{d}k_x\mathrm{d}k_y, 1 \leqslant i \leqslant 20$

Note: DN_{ij} represents the mean value of calibrated imagery intensity, A the imagery energy, $\langle\rangle$ the average operation, $\bar{P}$ the normalized SAR imagery energy spectrum, $\overline{h_i}$ the orthogonal function, and k_x and k_y the wave-number of the range and the azimuth direction, respectively.

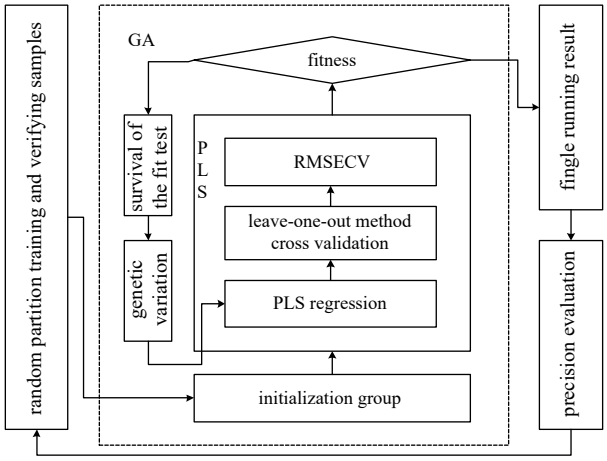

Fig. 3. The flow chart of SWH retrieval using the GA-PLS model.

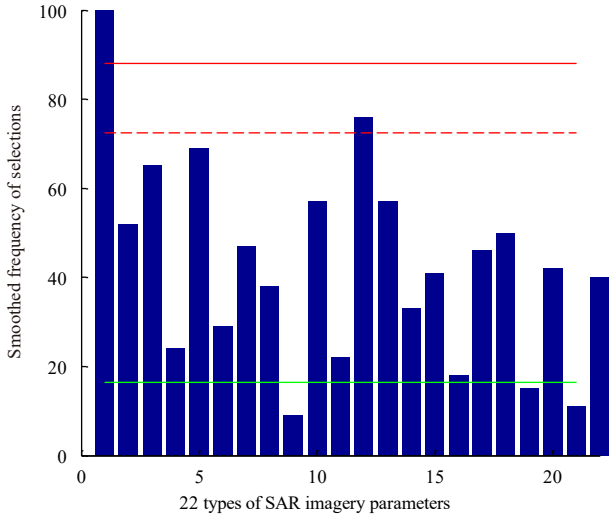

Fig. 4. Smoothed frequency of selections for the 22 types of SAR imagery parameters used in the emprical relationship.

previous runs. In Fig. 4, the red solid line represents the criterion of the optimal number of components, the red dashed line represents the criterion of the suggested model according to the cross validation, and the green solid line represents the non-significant components according to an F test. Through the GA-PLS modeling, the final retrieval model can be simplified as a linear regression equation of 2 types of components, which is described as follows:

$$y = a_1x_1 + a_2x_2 + \ldots, \tag{6}$$

where x_i is the polarimetric parameters selected from the GA-PLS model and a_i is the corresponding regression coefficient (i=1, 2, ...). According to the statistics in Fig. 4, x_i represents the backscattering coefficients σ_0 and S_{10}, and corresponding a_i is the normalized regression coefficients of –4.722 0 and –1.430 0.

It is noted that the orthogonal function $\overline{h_i}$ is composed of Gegenbauer polynomials and harmonic functions as follows:

$$h_{ij}(\alpha_k, \alpha_\varphi) = \eta(k_x, k_y)\, g_i(\alpha_k) f_j(\alpha_\varphi), 1 \leqslant i \leqslant 4, 1 \leqslant j \leqslant 5, \tag{7}$$

$$g_{n_k} = \sqrt{\frac{n_k + 3/2}{(n_k+2)(n_k+1)}}\, C_{n_k}^{3/2} \sqrt{1-\alpha_k^2}, \tag{8}$$

$$f_{n_\varphi - 1}(\alpha_\varphi) = \sqrt{\frac{2}{\pi}} \sin((\alpha_\varphi - 1)\alpha_\varphi), \tag{9}$$

$$f_{n_\varphi}(\alpha_\varphi) = \sqrt{\frac{2}{\pi}} \cos((\alpha_\varphi - 1)\alpha_\varphi), \tag{10}$$

where $g_i(\alpha_k)$ and $f_j(\alpha_\varphi)$ are the Gegenbauer polynomials and the harmonic functions, respectively. $\eta(k_x, k_y)$ is the weight function about the wavenumber in the respective range and azimuth in domain A, which is an elliptical integration area. $g_i(\alpha_k)$ is the general expression of Gegenberg polynomials, and in this paper n_k=1, 2, 3, 4. $f_{n_\varphi - 1}$ and f_{n_φ} correspond to the odd and even terms of the harmonic function, respectively.

After the building of our simplified relationship based on the GA-PLS model, we used another 42 774 scenes of SAR wave mode imagery to verify our empirical model for SWH retrieval. Figure 5 shows that a preferable retrieval results is acquired from the available matched datasets. Compared with the matched analysis SWH data, the assessment index of the correlation coefficient, the bias, the RMSECV and the scattering index (SI) are 0.78, 0.07 m, 0.76 m and 0.5, respectively. However, the model's precision deviates more with the increase of the SWH, especially for high sea states (SWH>5 m).

3.2 *Comparison with Jason-2 radar alimeter*

In addition to the precision validation by the ECMWF reanalysis of the SWH, we also make a comparison of the retrieved SWH using the Jason-2 radar altimeter data. Because of the limited spatial resolution, only 2 057 scene matched datasets are collocated with a spatial window of 100 km and a temporal window of 2 h. Figure 6a shows the correlation relationship between our empirical model and the Jason-2 measured SWH. There is a similar condition with Fig. 5a in which major bias appears in high sea states (SWH>5 m). It is shown in Fig. 5 that the assessment index of the correlation coefficient, the bias, the RMSECV and the SI are 0.76, 0.90 m, 0.04 m and 0.07, respectively.

3.3 *Comparison with CWAVE-ENV algorithm*

Moreover, we also make a comparison of the retreival SWH between our model and the well-known CWAVE_ENV model (Fig. 7). The result of Fig. 7a also shows a similar curve with the above validation, especially for Fig. 5a and Fig. 6a. Considering the simplified model, reliable results are acquired in low-moderate sea states, and major bias appears in high sea states (SWH>5 m). In addition, the coefficients of the CWAVE_ENV model and our simpified model are shown in the Appendix A and Appendix B.

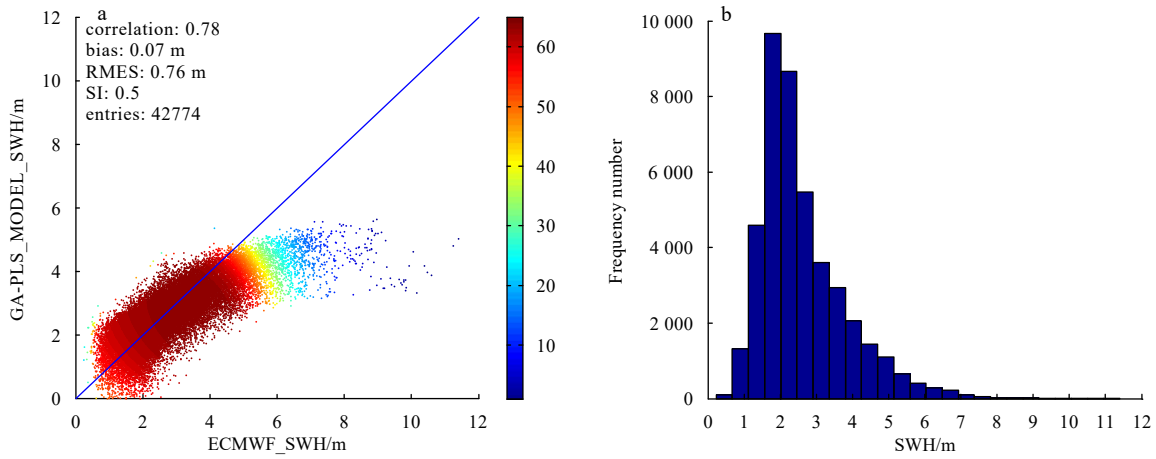

Fig. 5. Scatter plots of verification samples between the retrieved SWH of our simplified selected features and the SWH of the ECMWF reanalysis data. The colorbar represents the density of scattering points (a) and the histogram of the SWH (b).

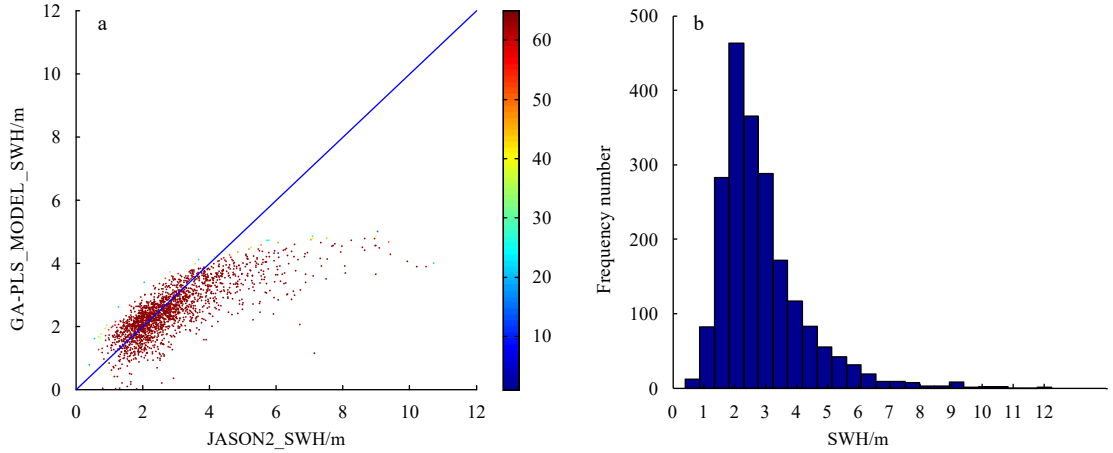

Fig. 6. Scatter plots of the verification samples between the retrieved SWH of our simplified selected features and the SWH of JASON-2. The colorbar represents the density of scattering points (a) and the histogram of SWH (b).

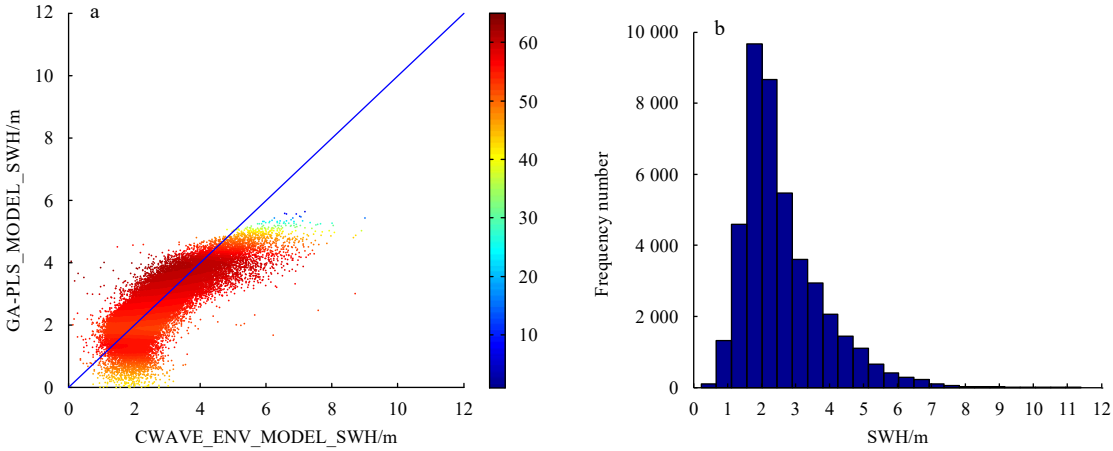

Fig. 7. Scatter plots of the verification samples between the retrieved SWH of our simplified selected features and the SWH of the CWAVE_ENV model. The colorbar represents the density of the scattering points (a) and the histogram of SWH (b).

Considering the validation performance of above three datasets, it is applicable to retrieve the SWH with our simplified model in low-moderate sea states. However, major bias appears in high sea states (especially for SWH>5 m). The major bias is probably due to the limited training samples of high sea states, which affect the model's building. Therefore, more datasets of high sea states will be collected in the next step for the model's robustness.

4 Discussion

To explain the large bias in high sea states (>5 m), the scatter plot for the comparison of the CWAVE_ENV model and the ECMEF data and the two histograms of the SWH for the training and verification datasets are presented in Figs 8 and 9, respectively. The assessment index of the correlation coefficient, the bias, the RMSECV and the SI are 0.92, 0.10 m, 0.43 m and 0.11, respectively. Although the retrieval result of the CWAVE_ENV performs well overall, large bias also appears in high sea states, especially for SWH>5 m. It is noticed that the color bar of Fig. 8 represents the density of scattering points, which means few samples are used for comparison. Figures 9a and b also reveal that both the training and verification samples lack high sea state samples. Generally, the high sea state always exists together with typhoon and other extreme weather. Therefore, we would like to collect more high sea state data to improve our SWH retrieval model for typhoon monitoring in a future study.

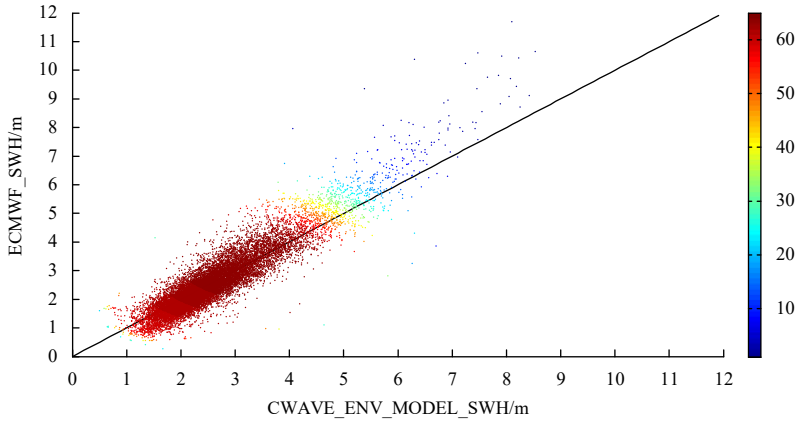

Fig. 8. Scatter plots of verification samples between the SWH of the CWAVE_ENV model and the SWH of the ECMWF. The color bar represents the density of scattering points.

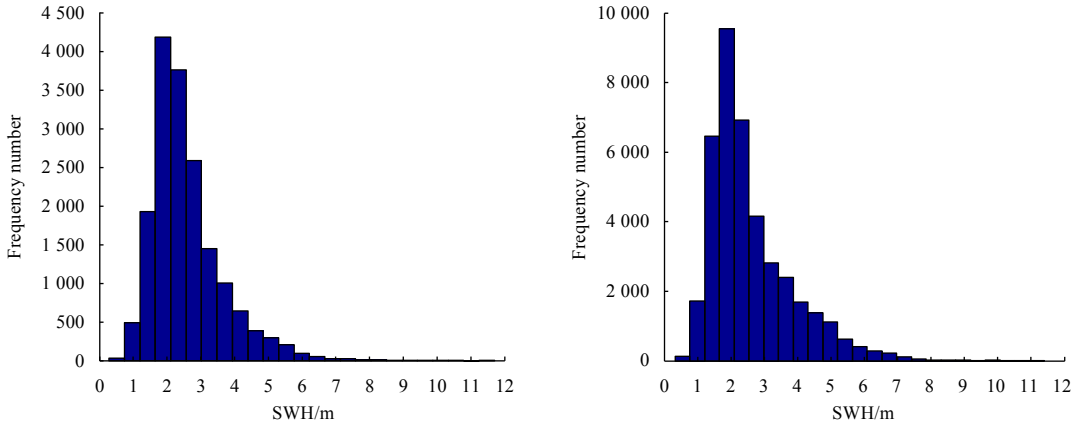

Fig. 9. Histogram of SWH. a. Training samples and b. verification samples.

5 Conclusions

This paper presents a simplified relationship between the SAR imagery parameters and the SWH. A total of 17 200 scenes of Envisat ASAR wave mode imagery and collocated ERA-Interim reanalysis data are used to build our model based on the GA-PLS model. This paper also further compares the SWH retrieval precision between our simplified model and the ERA-interim, the Jason-2 radar altimeter and the well-known CWAVE-ENV algorithm. The main conclusions of this paper are as follows.

After feature selection from the GA-PLS model, the results show that σ_0 and S_{10} make major contributions to SWH retrieval. The correlation coefficient, the bias, the RMSECV and the SI, respectively, are 0.78, 0.07 m, 0.76 m and 0.5 between our simplified model and the ERA-Interim reanalysis SWH. The assessment of the above indexes reveals the promise of our simplified model for ocean SWH retrieval. The comparison of our simplified model with the Jason-2 radar altimeter and the well-known CWAVE-ENV algorithm also shows satisfying retrieval results in low-mediate sea states. However, large bias appears in high sea states (especially for SWH>5 m).

Acknowledgements

We gratefully acknowledge the ESA for providing the Envisat ASAR wave mode data, and the ECMWF for providing the ERA-Interim reanalysis data.

References

Bruck M, Lehner S. 2015. TerraSAR-X/TanDEM-X Sea state measurements using the XWAVE algorithm. International Journal of Remote Sensing, 36(15): 3890–3912, doi: 10.1080/01431161.2015.1051630

Collard F, Ardhuin F, Chapron B. 2005. Extraction of coastal ocean

wave fields from SAR images. IEEE Journal of Oceanic Engineering, 30(3): 526–533, doi: 10.1109/JOE.2005.857503
Engen G, Johnsen H. 1995. SAR-ocean wave inversion using image cross spectra. IEEE Transactions on Geoscience and Remote Sensing, 33(4): 1047–1056, doi: 10.1109/36.406690
Geladi P, Kowalski B P. 1986. Partial least-squares regression: a tutorial. Analytica Chimica Acta, 185: 1–17, doi: 10.1016/0003-2670(86)80028-9
Gonzalez F I, Beal R C, Brown W E, et al. 1979. Seasat synthetic aperture radar: ocean wave detection capabilities. Science, 204(4400): 1418–1421, doi: 10.1126/science.204.4400.1418
Grieco G, Lin Wenming, Migliaccio M, et al. 2016. Dependency of the sentinel-1 azimuth wavelength cut-off on significant wave height and wind speed. International Journal of Remote Sensing, 37(21): 5086–5104, doi: 10.1080/01431161.2016.1226525
Hasselmann K, Hasselmann S. 1991. On the nonlinear mapping of an ocean wave spectrum into a synthetic aperture radar image spectrum and its inversion. Journal of Geophysical Research: Oceans, 96(C6): 10713–10729, doi: 10.1029/91JC00302
He Yijun, Perrie W, Xie Tao, et al. 2004. Ocean wave spectra from a linear polarimetric SAR. IEEE Transactions on Geoscience and Remote Sensing, 42(11): 2623–2631, doi: 10.1109/TGRS.2004.836813
Kawamura K, Watanabe N, Sakanoue S, et al. 2010. Testing genetic algorithm as a tool to select relevant wavebands from field hyperspectral data for estimating pasture mass and quality in a mixed sown pasture using partial least squares regression. Grassland Science, 56(4): 205–216, doi: 10.1111/grs.2010.56.issue-4
Leardi R, González A L. 1998. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. Chemometrics and Intelligent Laboratory Systems, 41(2): 195–207, doi: 10.1016/S0169-7439(98)00051-3
Li Bingyan, Gong Huaze, Shao Yun, et al. 2016. The consistency between Na content distribution at the subsurface and the lake body's movement in Lop Nur. International Journal of Digital Earth, 9(7): 662–675, doi: 10.1080/17538947.2015.1112439
Li Xiaoming, Lehner S, Bruns T. 2011. Ocean wave integral parameter measurements using envisat ASAR wave mode data. IEEE Transactions on Geoscience and Remote Sensing, 49(1): 155–174, doi: 10.1109/TGRS.2010.2052364
Li Lin, Ustin S L, Riano D. 2007. Retrieval of fresh leaf fuel moisture content using genetic algorithm partial least squares (GA-PLS) modeling. IEEE Geoscience and Remote Sensing Letters, 4(2): 216–220, doi: 10.1109/LGRS.2006.888847
Long R B. 1980. The statistical evaluation of directional spectrum estimates derived from pitch/roll buoy data. Journal of Physical Oceanography, 10(6): 944–952, doi: 10.1175/1520-0485(1980)010<0944:TSEODS>2.0.CO;2
Longuet-Higgins M S, Cartwright D E, Smith N D. 1963. Observations of the directional spectrum of sea waves using the motions of a floating buoy. In: Ocean Wave Spectrum. Englewood Cliffs, NJ: Prentice-Hall Inc, 111–136
Mastenbroek C, de Valk C F. 2000. A semiparametric algorithm to retrieve ocean wave spectra from synthetic aperture radar. Journal of Geophysical Research: Oceans, 105(C2): 3497–3516, doi: 10.1029/1999JC900282
Monaldo F. 1988. Expected differences between buoy and radar altimeter estimates of wind speed and significant wave height and their implications on buoy-altimeter comparisons. Journal of Geophysical Research: Oceans, 93(C3): 2285–2302, doi: 10.1029/JC093iC03p02285
Pleskachevsky A L, Rosenthal W, Lehner S. 2016. Meteo-marine parameters for highly variable environment in coastal regions from satellite radar images. ISPRS Journal of Photogrammetry and Remote Sensing, 119: 464–484, doi: 10.1016/j.isprsjprs.2016.02.001
Schulz-Stellenfleth J, König T, Lehner S. 2007. An empirical approach for the retrieval of integral ocean wave parameters from synthetic aperture radar data. Journal of Geophysical Research: Oceans, 112(C3): C03019
Shao Weizeng, Wang Jing, Li Xiaofeng, et al. 2017. An empirical algorithm for wave retrieval from Co-polarization X-band SAR imagery. Remote Sensing, 9(7): 711, doi: 10.3390/rs9070711
Shao Weizeng, Zheng Zhang, Li Xiaofeng, et al. 2016. Ocean wave parameters retrieval from sentinel-1 SAR imagery. Remote Sensing, 8(9): 707, doi: 10.3390/rs8090707
Song Kaishan, Li Lin, Li Shuai, et al. 2012. Hyperspectral retrieval of phycocyanin in potable water sources using genetic algorithm-partial least squares (GA-PLS) modeling. International Journal of Applied Earth Observation and Geoinformation, 18: 368–385, doi: 10.1016/j.jag.2012.03.013
Stopa J E, Mouche A. 2017. Significant wave heights from sentinel - 1 SAR: Validation and applications. Journal of Geophysical Research: Oceans, 122(3): 1827–1848, doi: 10.1002/2016JC012364
Wan Yong, Zhang Jie, Meng Junmin, et al. 2015. Exploitable wave energy assessment based on ERA-Interim reanalysis data-a case study in the East China Sea and the South China Sea. Acta Oceanologica Sinica, 34(9): 143–155, doi: 10.1007/s13131-015-0641-8
Wang He, Mouche A, Husson R, et al. 2016. A global distribution of crossing swell from Envisat ASAR Wave Mode data based on swell propagation. In: Proceedings of 2016 IEEE International Geoscience and Remote Sensing Symposium. Beijing, China: IEEE, 4657–4660
Wang He, Zhu Jianhua, Yang Jingsong. 2014. Error Analysis on ESA's envisat ASAR wave mode significant wave height retrievals using triple collocation model. Remote Sensing, 6(12): 12217–12233, doi: 10.3390/rs61212217
Wang He, Zhu Jianhua, Yang Jingsong, et al. 2012. A semiempirical algorithm for SAR wave height retrieval and its validation using Envisat ASAR wave mode data. Acta Oceanologica Sinica, 31(3): 59–66, doi: 10.1007/s13131-012-0206-z

Appendix A: Coefficient used in CWAVE_ENV model

Considering the interaction effect of each imagery parameter, the second order linear model is selected for model building.

$$W = a_0 + \sum_{i=1}^{N} a_i s_i + \sum_{i=1}^{N} \sum_{j=1}^{i} a_{i,j} \ \ s_i s_j, \tag{A1}$$

where N is the number of samples, W is the known SWH results, s_i is the imagery parameter, and a_i is the multi-variant coefficient of vector A. To confirm the regression coefficients of the CWAVE_ENV model, 22 types of imagery parameters are performed into cost function $J(A)$, which can be written as follows:

$$J_{\text{cost}}(A) = \sum_{j=1}^{N} \left(W^{(j)} - \sum_{i=1}^{k} A_i S_i^j \right)^2. \tag{A2}$$

Using a stepwise regression, the coefficient vector A can be proposed. It is noted that the multi-variant coefficients a_i of vector A are derived in Table A1.

Table A1. Multi-variant coefficients of vector A in the CWAVE_ENV model

Coefficient index	Value	Coefficient index	Value	Coefficient index	Value	Coefficient index	Value
1	0.874 2	13	–0.005 3	25	0.015 2	37	–0.018 7
2	–0.099 9	14	–0.017 8	26	0.062 2	38	–0.014 4
3	0.071 22	15	–0.003 4	27	–0.034 7	39	–0.021 9
4	0.031 9	16	0.027 6	28	0.035 5	40	0.054 3
5	–0.113 8	17	0.013 8	29	0.023 0	41	–0.014 2
6	0.007 0	18	–0.329 4	30	–0.008 5	42	0.017 2
7	–0.360 9	19	0.035 8	31	–0.015 3	43	–0.010 1
8	–0.024 2	20	0.034 2	32	–0.179 9	44	–0.004 9
9	0.038 6	21	0.080 4	33	0.043 8	45	0.017 1
10	0.004 4	22	–0.009 7	34	0.024 9	46	0.017 3
11	–0.006 4	23	0.015 8	35	0.026 3		
12	0.108 7	24	–0.004 7	36	0.035 6		

Appendix B: Coefficient used in the simplified model

In the GA-PLS model, the chromosome and fitness in the species correspond to a set of variables and the internal prediction of the derived PLS model, respectively. To avoid overfitting, the program implemented the following features:

(1) the parameters were set with the highest elitism,

(2) the model was set to be determined after 100 independent, short GA runs, and

(3) the frequency of the selection of the variables in each run was set to be a weighted average between the frequency for the selection of the variables in the starting run and the previous run.

The fitness function in which the individuals were subject to evaluation was the percentage of the predicted variance of a constituent abundance, which was defined as follows:

$$100 - \left\{ \left[\sum_{i=1}^{n} (y_i - y_i')^2 \Big/ n \right] \Big/ \left[\sum_{i=1}^{n} (y_i - y_i')^2 \Big/ k \right] \right\} \times 100, \tag{B1}$$

where y_i and y_i' are the measured and predicted values, n is the number of samples to be considered, and k=n–1 in the case of cross-validation.

In this study, the GA-PLS program was written in m-file and was compiled using a Matlab 6.1 compiler running an operating system on a Pentium IV with 256 M RAM.