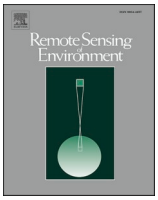
Remote Sensing of Environment 280 (2022) 113200

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse



Multiparametric sea state fields from synthetic aperture radar for maritime situational awareness

Andrey Pleskachevsky *, Björn Tings, Stefan Wiehle, James Imber, Sven Jacobsen

DLR, Maritime Safety and Security Lab Bremen, Am Fallturm 9, 28359 Bremen, Germany

ARTICLE INFO

Edited by Menghua Wang

Keywords:
SAR
Sea state parameters
Wave height
Wave period
Near real time services NRT
Machine learning
Sea state processor SSP
Sentinel-1
TerraSAR-X
Climate change initiative CCI

ABSTRACT

This paper introduces a method for estimating a series of sea state parameters from satellite-borne synthetic aperture radar (SAR). The method was realized in a near real time (NRT) application which allows for the processing of data from different satellites and modes. The algorithm estimates the total significant wave height H_s, dominant and secondary swell and windsea wave heights, first, and second moment wave periods, the mean wave period and the period of wind sea. The algorithm was applied to the Sentinel-1 (S1) C-band Interferometric Wide Swath Mode (IW), Extra Wide (EW) and Wave Mode (WV) Level-1 (L1) products and also extended to X-band TerraSAR-X (TS-X) StripMap (SM) products. The scenes are processed in raster and result in continuous sea state fields, with the exception of S1 WV, where averaged values for sea state parameters for along-orbit imagettes of 20 km × 20 km are presented.

The developed empirical algorithm consists of two parts: a first CWAVE_EX (extended CWAVE) part, based on a linear regression approach, and a subsequent machine learning part using the support vector machine (SVM) technique. A series of new data preparation steps (i.e. filtering, denoising) and new features estimated from SAR images are also introduced. The algorithm was tuned and validated using two independent global wave model hindcasts, WaveWatch-3 and MFWAM as well as National Data Buoy Center (NDBC) measurements. The achieved root mean squared errors (RMSE) for CWAVE_EX for the total H_s are 0.60 m for low-resolution modes S1 IW (10 m pixel) and EW (40 m pixel) and 0.35 m for S1 WV and TS-X SM (pixel spacing ca.1–4 m) in comparison to model predictions. The RMSEs of the retrieved wave periods are in the range of 0.45–0.90 s for all of the satellites and models considered. Similarly, the dominant and secondary swell, and wind sea wave height RMSEs are in the range of 0.35–0.80 m. The SVM postprocessing improves the accuracy of the initial results of CWAVE_EX for H_s and reaches an RMSE of 0.25 m for S1 WV. Comparisons to 64 NDBC buoys, collocated at distances shorter than 50 km to S1 WV imagettes worldwide, result in an RMSE of 0.41 m. All results and the methods presented are novel in terms of the accuracy achieved, combining the classical approach with machine learning techniques, and performing an automatic NRT processing of multiparametric sea state fields from L1 data with automatic switching for different satellites and modes.

The complete archive of S1 WV L1 Single Look Complex products from December 2014 until February 2021 was processed to create a sea state parameter database and validated using model hindcast and buoy measurements. The derived parameters are available to the public within the scope of the European Space Agency's Climate Change Initiative.

1. Introduction

This paper introduces an algorithm and processor for meteo-marine parameter estimation for near real time (NRT) applications. The paper is structured as follows: in the current Section 1, general background and methodology are reviewed. Section 2 is dedicated to the data used and Section 3 deals with the features extracted from radar images. Section 4 presents the results and cross-validation by applying linear regression. In Section 5 further algorithm improvement using machine learning is elaborated. The summary in Section 6 finalises the paper.

1.1. Applications of SAR imagery for meteo-marine parameters

In recent years, the ongoing development of spaceborne synthetic

* Corresponding author.
E-mail address: Andrey.Pleskachevsky@dlr.de (A. Pleskachevsky).

https://doi.org/10.1016/j.rse.2022.113200
Received 20 September 2021; Received in revised form 24 July 2022; Accepted 25 July 2022

aperture radar (SAR) together with corresponding data transfer and data processing infrastructures has made a series of oceanographic applications possible in near real time (NRT) (Schwarz et al., 2015; Pleskachevsky et al., 2016; Pleskachevsky et al., 2019). Several minutes after image acquisition, the acquired scene can be processed and all derived meteo-marine parameters are transferred to weather services that can use these products for forecast validations and to support marine traffic. In turn, sea state or sea ice information, for instance, can also be sent directly to a ship's bridge in order to optimise the ship routing (Frost et al., 2018). Oceanographic products, such as significant wave height and surface wind speed fields, ice coverage maps, oil spill locations, etc., can be processed in parallel from the same satellite image or from different satellites (Lehner et al., 2013) and combined to support Maritime Situational Awareness (MSA) by fusing data from various sources based on remote sensing, in situ measurements, forecast modelling and communication systems (Voinov et al., 2020). One essential data source for MSA is weather and sea state information with high spatial resolution, which can be measured only by remote sensing in most areas of the globe, especially in the open ocean and in polar regions. Furthermore, the fusion of oceanographic parameters not only generates additional value for MSA, but can also be applied for forecast model assimilation and to gain insights into the underlying physics. For instance, ship motion information from the Automatic Identification System (AIS) and background sea state information retrieved from SAR were combined in ship detectability models to connect the probability of SAR ship detection with the actual wind and sea state conditions (Tings et al., 2018a, 2018b, 2019; Tings, 2021).

1.2. Approaches to estimating maritime parameters from SAR

Traditionally, for estimating the sea state parameters from SAR imagery, the scenes or subscenes were converted into an image spectrum using, for example, the Fast Fourier Transformation (FFT). Two basic classical approaches have been considered for sea state parameter estimation:

- Transfer functions, which convert the image spectrum into a wave spectrum with a subsequent estimation of the integrated sea state parameters (e.g. Alpers and Rufenach, 1979; Hasselmann et al., 1996)).
- Empirical functions, which estimate the sea state parameters directly from features derived from the SAR scene, including parameters estimated from the image spectrum without its transformation into a wave spectrum (e.g. the CWAVE approach by Schulz-Stellenfleth et al., 2007; Lehner et al., 2012; Bruck, 2015).

The advantage of the first methods is that the resulting wave spectrum can be assimilated into a forecast system. Such assimilations have been practiced for the past decades (e.g. Abdalla et al., 2006, 2010). Nevertheless, for meaningful spectral transformations only acquisitions where pronounced wave patterns are imaged in SAR scenes can be used. This is usually associated with wavelengths exceeding the so-called *cutoff* for SAR wave imaging (Alpers and Rufenach, 1979; Holt, 2004). This method is particularly well suited to the open ocean with dominant long swell waves. However, in coastal areas, where most of Sentinel-1 (S1) Interferometric Wide Swath Mode (IW) scenes are acquired, this method can only be applied to about 30% of all scenes. Firstly, the waves are often too short in these areas and fall under the *cutoff*. Secondly, the SAR signal from the ocean waves is frequently distorted by a series of artefacts (e.g. ships, ship wakes, oil spills, etc.). All of these distortions and conditions result in a very noisy SAR image spectrum where wave-generated peaks are practically undiscernible. However, for estimating the sea state integrated parameters, the empirical functions are applicable to around 99% of all S1 IW as the image spectrum is complemented by a series of additional SAR features including texture analysis, e.g. grey level co-occurrence matrix (Pleskachevsky et al., 2019). The empirical approach allows for fast and robust processing of SAR scenes in NRT. A series of empirical algorithms for retrieving sea state parameters from SAR imagery have been constantly improved over the past few years (e.g. Stopa and Mouche, 2017; Rikka et al., 2018; Li and Huang, 2020).

Since the launch of the Sentinel-1 satellites in 2014 and 2016 (ESA, 2022b, Sentinel-1), a series of investigations for S1 SAR imagery have been carried out worldwide for different S1 modes. For example, an iterative nonlinear algorithm to estimate phase-resolved deterministic maps of wave-induced orbital velocities, from which elevation spectra were derived over ice-covered regions for S1 Wave Mode (WV), was demonstrated by Ardhuin et al., 2016. In Sun et al., 2018 a data assimilation for a spectral wave model using the S1 WV is presented. A semi-empirical algorithm for H_s and mean wave period retrieval from S1 StripMap were also reported by Shao et al., 2016.

In the past few years, machine learning techniques have taken a leading position in science, as their results are superior to those of analytical and simple empirical solutions when sufficiently large databases are available. Even though in their early stages, machine learning did not noticeably provide more accurate solutions than the classical approaches, today they already exceed them: in 2017, the accuracy of H_s obtained by applying neural networks (NN) in comparison to a traditional CWAVE method had not significantly improved the RMSE of ca. 0.50 m for S1 WV (Stopa and Mouche, 2017), whereas by using a deep learning technique in 2020 (Quach et al., 2020) the accuracy had significantly been improved to an RMSE of around 0.30 m.

The methods presented above are mostly concentrated on different SAR modes with different resolution for estimating individual parameters with a focus on significant wave height. In individual publications, in addition to wave height, the wave periods are estimated. In this paper an extended method, applicable for different modes and estimating multiparametric sea state fields reflecting a series of integrated sea state parameters has been considered.

1.3. Algorithm definition, specific points and objective of the study

Commonly, the estimation of a parameter from data is based on model functions which are only a part of the processing procedure. However, even the most optimal function will not be able to provide high accuracy for all cases without additional operations dedicated to data preparation such as filtering and denoising and control of results. In this paper, the term “algorithm” means the complete processing chain, with a series of steps each needed to reach high accuracy:

- data (subscenes) preparation and denoising: pre-filtering image artefacts, resampling, smoothing, etc.
- SAR features estimation and control
- empirical model function (EMF) for estimation of sea state parameters from SAR features
- control of results using filtering procedures

It is important to note that, although the algorithms are developed within scientific studies, they are aimed at practical applications. In this regard, some points impose boundary conditions for the algorithm's development. For algorithm applicability, not only the accuracy in terms of root mean square error (RMSE) and bias (BIAS) is to be considered, but also the percentage of rejected values and the processing speed are optimization parameters. In general, the uncertainty of measurements increases with the magnitude of the measured variable (e.g. H_s). This noise represents the distribution of the error across the variable domain and results not only from the inherent accuracy of the SAR methods, but is also related to a growing uncertainty in ground truth data (i.e. model or buoy) especially under storm conditions. Filtering out the ambiguous data on the one hand helps to improve the total accuracy, but on the other hand this operation increases the percentage of non-valid data. Therefore, not only the model function's accuracy in terms of RMSE, but

also filtering and outlier detection must be optimized.

The objective of this study was the development and technical realization of an algorithm to estimate a series of integrated sea state parameters under the following conditions:

- Fast automatic processing to allow working on massive archives of original Level-1 (L1) satellite SAR data with high-accuracy
- Uniform modular architecture to make the algorithm applicable to various satellites and modes
- Applicability for Near Real Time services by simple integration of the algorithm into the existing Sea State Processor (SSP) infrastructure.

While each step of the algorithm was developed in a series of iterations with testing different approaches, filtering and denoising procedures, SAR-features, etc., the material is presented in the final version implemented in an operational near real time service. The developments consist of two significant parts: a first part, based on a linear regression approach with tuning of coefficients by an analytical solution, and a subsequent machine learning part with model trainings. As the amount of data used are on the order of millions of samples, rounded values are given in the text; the exact values can be found in the Appendix in the corresponding tables Table A3 and Table A4.

2. Data and sea state parameters

In this work, first the ESA Sentinel satellites S1 were used. Then the improved method was adopted also for TerraSAR-X (TS-X). As ground truth, National Data Buoy Center (NDBC, 2022) buoy measurements and two independent hindcast models were applied.

2.1. SAR satellites and modes used

The Copernicus C-band satellites Sentinel 1-A (S1-A) and Sentinel 1-B (S1—B) operate at an altitude of 704 km with a ground speed of 6.8 km•s^{-1} and a radar wavelength of 5.6 cm, corresponding to a frequency of 5.4 GHz. The twin X-band satellites TerraSAR-X (TS-X, launched in 2007) and TanDEM-X (TD-X, launched in 2009) operate from a height of 514 km with a ground speed of 7 km s^{-1} and a radar wavelength of 3.1 cm, corresponding to a frequency of 9.6 GHz.

S1 WV (wave mode) acquires two parallel tracks with incidence angles of around 23° (wv1) and around 36° (wv2) with imagettes (small images with an approximate footprint of 20 km × 20 km, (ESA, 2022f, S1-WV) acquired every 200 km along each wv1 and wv2 track with a 100 km offset and distance of 100 km between wv1 and wv2 tracks. Using both wv1 and wv2 this means along-track imagettes each 100 km. The length of a track (relative orbit with ID) varies from around 1000 km (10 imagettes) to 12,000 km (120 imagettes). The nominal spatial pixel resolution of S1 WV is around 3 m depending on the local incidence angle. The scenes can be acquired in vertical (VV) or horizontal (HH) co-polarization. However, >95% of the data were acquired in VV polarization. In the context of this work, the method was designed and trained for calibrated S1 WV Single Look Complex (SLC) products (ESA, 2022c, manual). Each day, around 60 S1 WV products (ascending or descending tracks) each with around 120 individual imagettes for both S1-A and S1—B are acquired. From a technical point of view, the large file size of each individual S1 WV product of 2–10 GB in size (ca. 5 TB/month), each including hundreds of compressed individual imagettes, makes the downloading, unpacking and processing a real challenge and requires separate consideration.

S1 IW mode combines a large swath width with a moderate geometric resolution (ESA, 2022e, S1-IW). The individual IW images cover approximately 200 km in azimuth and 250 km in the range direction with a pixel spacing of 10 m. The original GRDH (Ground Range Detected High resolution) L1 products are available in single (HH or VV) or dual polarization (HH + HV or VV + VH). For sea state estimation, the VV or HH polarization data were used, with priority given to VV products.

S1 EW (Extra Wide) mode is similar to the IW mode, but the EW mode acquires data over a wider area than for IW mode using five subswaths. The EW mode acquires data over 400 km swath width with a coarser pixel spacing of 40 m (ESA, 2022d, S1-EW). Both IW and EW images with corresponding swaths of around 250 km and 400 km were mostly acquired in strips of up to 2000 km covering the ocean surface.

While WV imagettes and IW/EW images are taken a hundred times per day, the StripMap (SM) mode has been rarely acquired with only a few images per day and has not been considered in this study. Fig. 1 shows an example of the daily acquisitions by S1-A and S1—B for 2020-09-01. As can be seen, during one day the following images were acquired:

- 54 S1 WV tracks/orbits with 6480 individual imagettes,
- 894 S1 IW individual images,
- 229 S1 EW individual images,
- 10 S1 SM individual images

TS-X SM (StripMap) data used for this study are Multi-Look Ground Range Detected (MGD) standard products with ca. 30 km × 50 km swath size and ca. 3 m pixel spacing (ESA, 2022g, TS-X). The TS-X StripMap are mostly acquired in coastal areas. Compared to S1, the TS-X archive is rather limited with a few thousand acquisitions; each TS-X acquisition must be ordered by a user, there is no general acquisition schedule as for S1.

2.2. Wave model hindcast

As ground truth data, two independent spectral wave models were employed:

- WaveWatch-3 (WW3) model with a resolution of 0.5 degrees (NOAA, 2022), spatially interpolated on 0.25 degrees (ca. 20–25 km, which corresponds to the S1 WV imagette size, data available for the entire S1 mission).
- Météo-France WAve Model (MFWAM) runs (CMEMS, 2022; Ardhuin et al., 2010) with a spatial resolution of 1/12 degrees (data available from April 2016 onwards).

The model results, provided in 3-hourly steps, are temporally interpolated. The following model data filtering was applied:

- Land filtering: only data where at least one model grid point separates a SAR subscene from the land was used, which means the data points are around 25 km away from the land.
- Ice filtering: only model data for $-60° <$ latitude$< 60°$ were originally used in order to avoid ice coverage. This filtering corresponds to the setup used by Stopa and Mouche (2017) and was applied to the ESA Round-Robin competition dataset (see Section 2.5). After detailed investigations, the area south of −55° was also excluded due to high ice content.

2.3. Tuning, training and validation SAR data sets

All collected SAR data were divided into tuning, training and validation sets. Detailed information with data set names is given in Appendix Table A3 (S1 WV) and in Table A4 (S1 IW, S1 EW, TS-X SM). Considering different satellites, modes, polarization, training, validation, etc., a total of eleven data sets are used.

S1 WV. The S1 WV orbits collocated with NDBC buoys from 2015 to 2017 were applied for tuning using ground truth from WW3 and MFWAM (Table A3: **S1-WV_LR_TUN**), the 2018 orbits were used for validations (Table A3: **S1-WV_LR_VAL**). Additionally, the data from 2019 were used for an independent validation (RoundRobin; Table A3: **S1-WV_LR_RR)** for comparison of different algorithms (linear and

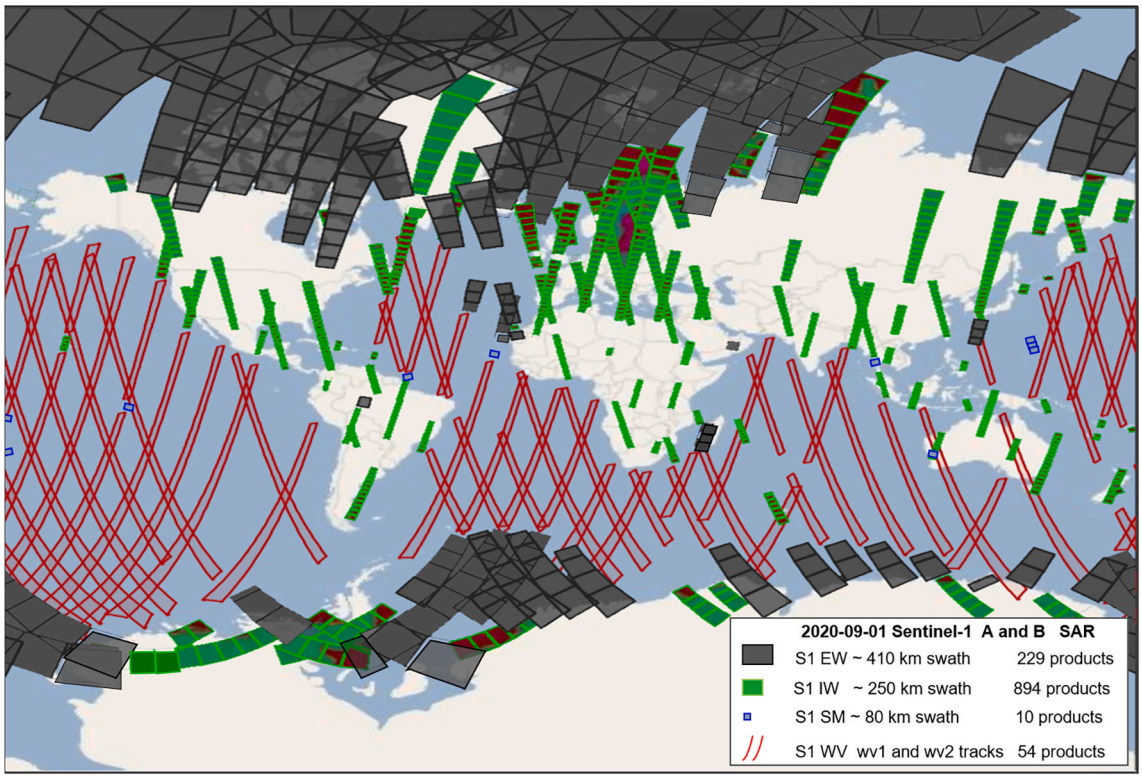


Fig. 1. An example of S1-A and S1—B worldwide acquisitions on 2020-09-01 (ESA). There are 54 S1 WV tracks (red), 894 S1 IW images (green), 229 S1 EW images (grey) and 10 StripMap images (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

machine learning approaches in this study). For 2015–2017, the number of samples is not large in comparison to the full data archive, as at the beginning of the studies the data were downloaded as individual products from the ESA DataHub considering only S1 WV orbits collocating with stationary buoys. Worldwide, 64 NDBC buoys were found to be collocated closer than 50 km to the border of the S1 WV (ca. 60 km to imagette center) with 61 S1 WV orbits which build the first tuning data set. The distance of 50 km was chosen as an optimal trade-off between having a larger number of collocations and maintaining acceptable local sea state variations (in fact, worldwide only 15 buoys are collocated closer than 20 km). Each S1 WV orbit/product can be acquired 2–3 times per month for both S1-A and S1—B (ca. 150 orbits/month from a total of ca. 2000 orbits/month). Later an automatic download of the S1 archive was organized for all S1 WV data. In this way, the machine learning data sets for training and validation are an order of magnitude larger (Table A3: **S1-WV_SVM_TR, S1-WV_SVM_VAL**).

S1 IW. A data pool with around 1700 S1 IW images was collected for the period from 2016 to 2020. The IW scenes were processed using a 10 km raster. Each IW image covers an area of around 250 km × 200 km, which results in around 500 samples. However, as many images are acquired in coastal areas and the relatively coarse grid resolution of the used models required an additional distance of around 15–20 km from the coastline (at least more than one grid point from land), only around 60% of all potential subscenes could be used for tuning/validation. The data cover different regions and sea state conditions including strong storms and hurricanes, for instance hurricane “Irma” in 2017. The data includes acquisitions in the North, Baltic, Black, and Mediterranean Seas, North Atlantic, the USA/Canada west and east coasts, including the Aleutian Islands and Hawaii, Australian coasts. Finally, around 520,000 samples were collected and randomly divided into a tuning data set with 300,000 samples (Table A4: **S1-IW_TUN**) and a validation set with around 220,000 samples (Table A4: **S1-IW_VAL**).

S1 EW. Similar to IW, EW images were processed with a raster. EW images are rarely acquired near the coast and more often taken in polar regions where sea ice is present. However, a large number of S1 EW images were still acquired in ice-free regions and even around the equator. To completely eliminate the influence of sea ice, the collocations for $-55° < \text{latitude} < 60°$ were used. The processing with a raster of 20 km resulted in around 400 collocated samples per EW image. Around 2000 S1 EW images were collected, which resulted in around 1.2 Mio samples. Similar to the IW mode, the data were divided into 800,000 tuning (Table A4: **S1-EW_TUN**) and around 360,00 validation samples (Table A4: **S1-EW_VAL**).

TS-X SM. For this study, the collected archive of the TerraSAR-X SM data with around 2000 scenes was used. The TS-X SM images were acquired in VV, HH and also in dual polarization (both HH and VV images). Around 60% of all pooled TS-X SM data had dual polarization. As the function was tuned for each polarization individually, the dual-polarization products allowed to extend the number of samples for each polarization. Each scene covers an area of around 30 km × 50 km, however, TS-X scenes are usually acquired near the coast and over harbours, so only around 20% of the images covered ocean surface. Processed with a raster step of 1.5 km, the collected TS-X SM images resulted in around 220,000 model collocations divided into 120,000 collocations for tuning and the rest for validation. Names of four TS-X datasets (tuning and validation for each VV and HH) see in Table A4.

2.4. *In-situ measurements*

Observations from in situ buoys were used as independent data for validations. Measurements from different providers were used with focus on NDBC, Environment and Climate Change Canada (ECCC, 2022) and European Marine Observation and Data Network (EMODNET, 2022). A detailed list can be seen in Pleskachevsky et al., 2019.

The stationary NDBC buoys provide mostly hourly measurements. However, only around 30% of all NDBC stations provide significant

wave height. Worldwide, 64 NDBC buoys were found to be collocated closer than 50 km to the border of the S1 WV orbits. For validations, the data for each buoy are temporally interpolated, data with a measurement time gap of over 6 h were excluded. Optimally, for S1 WV around 2000 collocations would be possible every year. However, in reality only around 1500 collocations per year can be pooled, as will become apparent from this work. There are two main sources for loss of potential collocations:

- Time gap between buoy measurements of above 6 h (or a buoy has failed for months). Some buoys were taken out of service (e.g. NDBC-43010 since 2018).
- S1 orbit was not completely acquired. This mostly concerns the ascending orbits longer than 12,000 km with >120 imagettes, starting in the Antarctic and collocating with buoys at the Aleutian Islands.

2.5. Sea state parameters considered, ground truth and round-Robin dataset

In this work, eight integrated sea state parameters are considered (Copernicus Resources, 2022; NOAA Manual, 2022).

In the study three different ground truth data sources were considered: MFWAM, WW3, and NDBC. An example for comparison of all three ground for location NDBC-46001 (56.304°N; 147.920°W) for February 2018 is shown in Appendix Fig.A1. The H_s comparison resulted into an RMSE = 0.26 m for MFWAM/NDBC and RMSE = 0.23 m for MFWAM/WW3 at NDBC buoy locations. Generally, in terms of H_s, the ground truth uncertainty can be assessed to around 0.25 m.

For the entire tuning, trainings and validations of H_s only MFWAM data were selected. This selection was a consequence of a cross-validation using three independent ground truth datasets built using WW3, MFWAM and mixed MFWAM/WW3 data: H_s estimated using the model function tuned with only MFWAM data resulted in higher accuracy than when tuned using WW3 and mixed MFWAM/WW3 datasets by comparison to the NDBC measurements. The details of this cross-validation will be presented in Section 4.3. However, not all parameters are included into the MFWAM results. The mean period Tm_0, first moment wave period Tm_1, and second moment period Tm_2 were tuned and validated with WW3 data.

In the scope of the ESA CCI Round-Robin (RR) competition, different algorithms were compared and a uniform ground truth validation RR dataset was created using MFWAM model data (Table A3: **S1-WV_RR**). This dataset includes MFWAM-H_s corresponding to imagettes from 61 S1 WV orbits collocated to NDBC buoys under 50 km distance for 2019; the collocated buoys-H_s are also collected in the RR data set with 2004 measurements.

3. Method and SAR features

The following chapter describes the processing method, filtering and denoising procedures and SAR features.

3.1. Sea state processor extensions

The Sea State Processor (SSP) for SAR data processing was developed originally for TerraSAR-X StripMap (Pleskachevsky et al., 2016) and later extended to the S1 IW mode (Pleskachevsky et al., 2019). The SSP was designed in a modular architecture for different satellite modes. The Ground Station Neustrelitz (German Aerospace Center DLR) applies the SSP as part of its NRT demonstrator service that involves a fully automated daily provision of surface wind and sea state parameters from S1 IW images in the North and Baltic Seas.

Within the scope of this work, the SSP was extended for S1 WV and S1 EW. A series of improvements in terms of accuracy, SAR image error corrections and processing speed were made (Pleskachevsky et al., 2022). Due to implemented parallelization, a fine raster can be processed: for example, a S1 IW image (~200 km × 250 km) can be processed with a 1 km raster (~50,000 subscenes) in one minute. The maritime environment and ship detection products are combined in layers (an example screen is given in Appendix Fig.A2).

The processing starts with SAR scene reading and calibration (ESA manual) followed by filtering of NRCS outliers. Then, depending on the SAR mode used, different kinds of smoothing and denoising are performed. As a next step, the relevant SAR features are estimated. After the features are extracted and controlled, the corresponding model functions (S1 WV, S1 IW, S1 EW, TS-X) are applied and then the results are controlled and stored. The detailed descriptions of the procedures are given in Pleskachevsky et al. (2019).

3.2. Subscenes preparation prefiltering and denoising

In a classic way, the estimation of sea state parameters is based on an NRCS analysis of subscenes. The main difference in processing the various SAR modes are:

- S1 IW, S1 EW and TS-X SM scenes were processed to sea state fields (raster). The raster step (distance between analysed subscenes) may differ.
- S1 WV imagettes are processed to one value for every along-track imagette (along-track points).

Direct application of the model function to a image features of a subscene can lead to inaccuracies in the range of meters in terms of H_s (outliers). The sources of these errors are both natural and man-made SAR image artefacts: ships and ship wakes, large wind farm constructions, current boundaries, wind streaks, oil and algae films, atmospheric fronts, etc. Thus, as a first step, the selected subscenes are pre-filtered for extreme values both significantly higher and lower than the average NRCS. To do this, a simple outlier detection applying a sliding window (sub-subscenes) of 150 m × 150 m for each subscene has been applied. The statistics for a subscene and its sub-subscenes identifies the pixels with extreme values in a sub-subscenes, which are then reset to the mean value of the subscene. The details can be found in Pleskachevsky et al., 2016, 2019.

S1 WV: For the WV mode, imagettes are processed initially in a raster, each subscene with 1024 × 1024 pixels covers an area of around 4.6 km × 4.6 km. Furthermore, an average value for each feature from valid subscenes using filtering based on min/max values for the features has been implemented (more details can be found in the next section "Features" and in Appendix, Table A1 "Criteria for validity of subscenes"). If all subscenes are recognized as non-valid, the complete imagette is marked as non-valid. After processing the entire WV archive, it could be concluded that around 18% of all imagettes have at least one non-valid subscene and ca. 1.2% of the imagettes were recognized as completely non-valid. For this work, different steps for the sliding window were tested, yielding a varying number of subscenes per imagette. For the final constellation, nine subscenes per imagette (3 × 3 = 9 for range×flight directions) was found to be optimal. A test with 25 subscenes per imagette (5 × 5 = 25) did not result in a better RMSE for H_s, but increased the processing time by about 15%.

S1 IW and S1 EW: After a series of tests with different setups, a denoising operation using smoothing for the subscenes was employed, significantly improving the results for these coarse modes. For an initial subscene of 256 × 256 pixels, a resampling with a factor of four (each pixel is divided into 4 × 4 = 16 pixels with the same NRCS value and with a size of 1/4 of the original spacing) followed by two-dimensional Gaussian smoothing has been applied. The resulting subscene is 1024 × 1024 pixels, whereby the modified pixel resolution is 2.5 m for IW (the subscene covers an area of 2560 m × 2560 m) and 10 m for EW (the subscene covers an area of 10,240 m × 10,240 m). By applying this operation, the improvement to RMSE~30 cm for H_s was reached

(Pleskachevsky et al., 2019). The next operation that improves the result is a smoothing for each SAR image estimated feature for each grid point with neighbor points to remove each feature's outliers. It was found out experimentally that for optimal smoothing, the scene should be processed with a raster step (distance between subscenes) between 1/4 and twice the subscene size.

TS-X SM/SL: For retuning the algorithm for TS-X SM data, all settings used for S1 IW and S1 EW excluding the resampling were applied.

3.3. SAR features systematization

All features estimated from a SAR subscene are summarized in Table 2 and separated into three groups, depending on their application and publication status:

Group-1: Features introduced in the original CWAVE (Schulz-Stellenfleth et al., 2007) and extended by Stopa and Mouche, 2017 (skewness, kurtosis).

Group-2: Features not used in the original CWAVE but already published and applied in other scientific branches, for example Grey Level Co-occurrence Matrices (GLCM) for ice coverage classification (e.g. Ressel et al., 2015) and for oil detection (e.g. Singha et al., 2013). These features are summarized in Pleskachevsky et al., 2019. In addition to these features, two further features used in this study and applied to the image spectrum are:

- Longuet-Higgins parameter P_{LH} represents spectral width (Longuet-Higgins, 1984).
- Goda peakedness parameter P_G represents spectrum narrowing (Goda, 1970).

P_{LH} and P_G parameters were introduced initially to evaluate the wave spectra and are based on combinations of different statistical moments of the image spectrum and integrated spectrum energy.

Group-3: New features introduced and tested within the framework of this study.

SAR features are also classified into five different types:

Type-1: NRCS and NRCS statistics (variance, skewness, kurtosis, etc.).

Type-2: Geophysical parameters (wind speed using CMOD-5 algorithms for C-band (Hersbach, 2008) and XMOD-2 for X-Band (Li and Lehner, 2014).

Type-3: GLCM parameters (entropy, correlation, homogeneity, contrast, dissimilarity, energy, etc.).

Type-4: Spectral parameters, based on image spectrum integration of different wavelength domains (0–30 m, 30–100 m, 100–400 m, etc.) and spectral width parameters (P_{LH} and P_G).

Type-5: Spectral parameters using products of normalized image spectrum with orthonormal functions and cutoff wavelength estimated using autocorrelation function (ACF).

3.4. Image spectrum and integration

One of the basic variables represents the SAR image spectrum $ISP(k_x, k_y)$ obtained using FFT applied to the radiometrically calibrated, filtered, land-masked and normalized subscenes with a size of 1024 × 1024 pixels in wave number domain k with a direction x for satellite range and y for flight direction as introduced in Pleskachevsky et al., 2019. For the estimation of CWAVE parameters, the normalized ISP is applied. The products of the normalized ISP with orthonormal functions then results in a series of features, 5.1–5.20 (Schulz-Stellenfleth et al., 2007) (Type-5, Group-1) in Table 2.

The whole 2-D image spectrum integration results into E_{ISP} and is limited by $k_x^{max} = k_y^{max} = 2\pi/L_{min}$ (rad/m), where L_{min} is the pixel spacing of the subscene and $k_x^{min} = k_y^{min} = 2\pi/L_{max}$, where L_{max} is the subscene size (the subscene resolution in x and y directions are equal).

Within the frame of this work, we named feature 4.3 after Wolfgang Rosenthal (RIP 2016), who suggested to introduce this parameter. The Rosenthal-parameter E_R indicates the image spectrum energy integrated with additional scaling of $1/k$, gained from k_x and k_y components as follows:

$$E_R = \iint_{k_x^{min} k_y^{min}}^{k_x^{max} k_y^{max}} \frac{ISP(k_x, k_y)}{\sqrt{k_x^2 + k_y^2}} dk_x dk_y \quad (1)$$

In contrast to a direct integration E_{ISP}, the E_R integration amplified the longer wavelength signals by $1/k$-weighting. In this way, objects such as ships or ship wakes in a subscene significantly increase the value of E_R and can thus be filtered out.

Note that using only a few basic features, such as NRCS, *nv* (Mean Intensity and Normalized variance, see Table 2), local surface wind U_{10}, and the integrated image spectrum energy E_{ISP}, are sufficient to estimate H_s with a low accuracy of around 80 cm. In this study, an additional 22 first-order features introduced in Schulz-Stellenfleth et al., 2007, three additional features introduced in Stopa and Mouche, 2017 (cutoff (Kerbaol and Chapron, 1998), skewness, kurtosis), as well as several newly introduced additional features are applied. Each addition slightly improves the resulting accuracy by 0.1–3.0 cm with an accumulative improvement of around 15 cm (ca. 30%) in comparison to the accuracy achieved by Stopa and Mouche, 2017 (RMSE~50 cm using CWAVE and WW3 hindcast).

Since all features are in fact connected, the RMSE improvement of each feature depends on the selected ensemble of features. The largest contribution can be gained using E^{100} and E^{400} (features 4.5, 4.6 in Table 2) by direct spectrum integration. E.g., using only E^{100}, in addition to established CWAVE, an improvement of ca. 5 cm can be reached. A comparable improvement can be reached also using E^{400}. However, when used together, the total improvement is only around 7 cm. Also, contrary to expectations, using λ_c (cutoff) instead of E^{100} or E^{400} brought only a weak improvement in the order of 1 cm.

3.5. Newly introduced SAR features

The new features tested in this study belong to two types (see Table 2). From all tested features, only those improving the results in term of H_s by at least 1 mm were applied in two subgroups:

Subgroup-A: features 1.7, 1.8, 1.9 (Type-1, NRCS and NRCS statistics).

Subgroup-B: features 4.15, 4.16 (Type-4, spectral).

The newly introduced features of type-1 (Subgroup-A) are based on using the Complementary Cumulative Distribution Function (CCDF) of NRCS in a subscene, which is equal to "1-CDF" (Cumulative Distribution Function). The parameter *CCDF* has been discretely calculated for the series of NRCS-bins while the number of NRCS values exceeding the predefined thresholds-bins (see Appendix Table A2) are accounted for. In this way, *CCDF* is a decreasing curve representing the cumulative probability distribution of the pixels' brightness. For example, in case of bin NRCS = 0, the value of the *CCDF* is equal to the number of all pixels in a subscene (NRCS≥0 for all pixels in a valid subscene). An example of a *CCDF* for three typical cases for low, middle, and high sea state can be seen in Fig.A3.

Using a weighting for the bins, amplifying the signal for high NRCS values, the parameters are introduced using an integration of the *CCDF* curve:

$$INT = \sum_{n=1}^{n_{bins}} CCDF_n \cdot G_n \quad (2)$$

$$INT\_LOG = \sum_{n=1}^{n_{bins}} log(CCDF_n) \cdot G_n \quad (3)$$

where G_n is a weighting factor of each NRCS-bin n, $CCDF_n$ represents the number of pixels with an NRCS exceeding the n-bin NRCS value, n_{bins}

= 21 is the total number of bins (see Table A4). The *INT_LOG* additional amplifies the contribution of high NRCS.

An additional simple parameter N_{HV} representing one bin of the *CCDF* for very high values of NRCS for outlier pixels (experimentally set to NRCS = 1500 for S1 wv1 and NRCS = 500 for S1 wv2) was also introduced (Type-4, subgroup-b). For low sea state N_{HV} is usually zero and generally grows with increased wave height.

The newly introduced features of Type-4 (Subgroup-b) are based on the cumulative projection curves ISP_{kx} and ISP_{ky} of the 2-D image spectrum on the k_x (range direction) and k_y (flight direction) axes, respectively:

$$\mathrm{ISP}_{\mathrm{kx_j}} = \sum_{\mathrm{i}=1}^{\mathrm{n_{max}}} \mathrm{ISP(j,i)}, \mathrm{ISP}_{\mathrm{ky_i}} = \sum_{\mathrm{j}=1}^{\mathrm{n_{max}}} \mathrm{ISP(j,i)} \tag{4}$$

where the $i = 1,2\ldots n_{max}$ and $j = 1,2\ldots n_{max}$ are the corresponding bin numbers of the 2-D image spectrum *ISP* in *x* (range) and *y* (flight) direction. The projections have the same dimension $n_{max} = 1024$ (size of FFT-box and spectrum for both *x* and *y* directions). The $dk_y = dk_x = const$ are reduced by integration. An example of both curves is shown in Appendix Fig.A4. Both curves are compared and combined in various ways:

REL- ratio between integrated square for each projection curve, weighted by actual k_x and k_y (Eq. 5).

S_{yx}- ratio between integrated squares for positive and negative differences of the curves taken for the same numbers $i = j$ ($k_x = k_y$) (Eq. 6).

CONV - convolution of both curves (Eq. 7)

$$\mathrm{REL} = \sum_{\mathrm{j}=1}^{\mathrm{n_{max}}} \left(1/\mathrm{k_{x_j}}\right) \mathrm{ISP}_{\mathrm{kx_j}} \Big/ \sum_{\mathrm{i}=1}^{\mathrm{n_{max}}} \left(1/\mathrm{k_{y_i}}\right) \mathrm{ISP}_{\mathrm{ky_i}} \tag{5}$$

$$\mathrm{S_{yx}} = \frac{\mathrm{DIFF^-}}{\mathrm{DIFF^+}} \text{ where } \begin{cases} \mathrm{DIFF^-} = \left| \sum_{\mathrm{ji}=1}^{\mathrm{n_{max}}} \left(\mathrm{ISP}_{\mathrm{kx_{ji}}} - \mathrm{ISP}_{\mathrm{ky_{ji}}} \right) \right| \text{ for } \mathrm{ISP}_{\mathrm{kx_{ji}}} < \mathrm{ISP}_{\mathrm{ky_{ji}}} \\ \mathrm{DIFF^+} = \sum_{\mathrm{ji}=1}^{\mathrm{n_{max}}} \left(\mathrm{ISP}_{\mathrm{kx_{ji}}} - \mathrm{ISP}_{\mathrm{yx_{ji}}} \right) \text{ for } \mathrm{ISP}_{\mathrm{kx_{ji}}} > \mathrm{ISP}_{\mathrm{ky_{ji}}} \end{cases} \tag{6}$$

$$\mathrm{CONV} = \mathrm{ISP_{k_y}} * \mathrm{ISP_{k_x}} \tag{7}$$

where $ji = 1,2\ldots n_{max}$ is the uniform numbering of k_x and k_y bins by comparisons of the curves. The idea behind introducing these parameters was to incorporate the nonlinearity of the SAR imaging mechanism, where the image spectrum reflects *cutoff* effects and differs in its shape from a wave spectrum. Under a stronger *cutoff* effect, the shape of the k_x and k_y image spectrum projections are more different (the *y*-curve has a stronger peak in comparison to the *x*-curve, see Fig.A4). However, where the image spectrum is close to the shape of a wave spectrum (more linear imaging), the shapes of both projections are more similar, and the features in Eq.5–7 are closer to 1.

The features selected for testing the validity of the subscenes are summarized in Table A1 "Criteria for subscenes validity".

4. Linear regression application

The following chapter deals with the implementation of the SAR features in the linear regression function, followed by validation for eight sea state parameters.

4.1. Applying the CWAVE linear regression, extended algorithm CWAVE_EX

In the original CWAVE approach (Schulz-Stellenfleth et al., 2007), the primary features extracted from the SAR image are supplemented with secondary features. The secondary features are combinations of primary features in quadratic form. In the present work, in addition to these multiplication products, the inverse features were also added, improving the results by around 2 cm for H_s in case of S1 WV. The number of all possible quadratic combinations is very high. However, in the final algorithm, only 77 secondary features (27 quadratic and 50 inverse) were added, each improving the resulting RMSE by at least 0.5 mm by addition of a feature to the already created feature-ensemble; the final feature matrix includes 131 features.

Since the SAR features differ by orders of magnitude, their use in a linear combination leads to an underestimation of the information for parameters with small values. After testing different normalization methods, the normalization using the mean (MEAN) and standard deviation (STD) was applied, improving the original CWAVE (uses partial normalization of individual feature components) by a further ca. 3 cm for H_s. In the work below, the model function CWAVE extended with the newly introduced features is denoted as CWAVE_EX (CWAVE Extended).

4.2. Cross validations

In order to define the optimal constellation of the ground truth data, a cross validation was carried out for H_s. The cross validations were completed using MFWAM, WW3, buoy data and S1 WV (Table A3: **S1-WV_TUN, S1-WV_VAL**). The model function was tuned independently using three setups with:

Setup-1: MFWAM collocated H_s used for tuning.

Setup-2: WW3 collocated H_s used for tuning.

Setup-3: MFWAM+WW3 collocated mixed H_s-dataset used for tuning.

At first, the direct validations were carried out with modelled H_s, i.e. MFWAM-tuned with MFWAM, WW3-tuned with WW3, MFWAM+WW3-tuned with MFWAM+WW3. Then, the coefficients for all three setups were swapped, so that the results gained by using the model function tuned with one ground truth were compared with another ground truth. Finally, all three variants were validated with NDBC buoys. Table 3 presents the results of this cross validation. Eight results (wv1/wv2) represent different combinations of different ground-truth-source for tuning (models) and different ground-truth for validation (models and NDBC). As can be seen, for buoy validation, an optimal RMSE = 42 cm is achieved when using only the MFWAM model data in the tuning step. The more accurate results gained using only MFWAM data can be explained by differences in the models' spatial resolutions: the WW3 spatial resolution is around 50 km and more than twice the dimensions of one S1 WV imagette covering around 20 km.

4.3. Results for linear regression function CWAVE_EX

The model function CWAVE_EX was tuned and validated for eight parameters (Table 1) and for four described satellite modes (datasets Table A3 and Table A4). The resulting total RMSE for all sea state conditions are summarized in Table 4.

Fig. 2 shows the CWAVE_EX validation with hindcast data (Table 1) for S1 WV mode for two of the most important parameters H_s and T_{m2}, used for a series of technical applications. The results for wv1 and wv2 imagettes are plotted separately. Fig. 3 presents the scatterplots for H_s and T_{m2} for S1 IW, S1 EW and TS-X. The scatterplots for other

Table 1
Sea state parameters considered and ground truth source.

N°	Parameter	Units	Symbol in text	Source
1	significant wave height	m	H_s	MFWAM
2	mean wave period	s	T_{m0}	WW3
3	first moment wave period	s	T_{m1}	WW3
4	second moment wave period	s	T_{m2}	WW3
5	wave height swell dominant system	m	$H_s^{swell\text{-}1}$	MFWAM
6	wave height swell secondary system	m	$H_s^{swell\text{-}2}$	MFWAM
7	significant wave height windsea	m	H_s^{wind}	MFWAM
8	mean period windsea	s	T_{wind}	MFWAM

Table 2
First-order SAR features.

Feature type	Feature description	Symbol	Feature group
1. NRCS and NRCS statistics	1.1. Mean Intensity of subscene scaled	*MI*	2
	1.2. STD of NRCS	*STD*	2
	1.3. Normalized variance	nv	1
	1.4. Variance of normalized NRCS	*Nv*	2
	1.5. Skewness	skew	1
	1.6. Kurtosis	kurt	1
	1.7. Accounting for high values of NRCS in subscene	N_{HV}	3
	1.8. Integral over NRCS CCDF distribution	*INT*	3
	1.9. Integral over NRCS CCDF distribution, logarithmic scale.	*INT_LOG*	3
2. Geophysical	2.1. Wind using CMOD5 (Sentinel-1) and XMOD2 (TerraSAR-X)	U_{10}	1
3. GLCM, (grey level co-occurrence matrix) feature analysis	3.1. GLCM-mean	*GLCMM*	2
	3.2. GLCM-variance	*VAR*	2
	3.3. GLCM-entropy	*ENTROPY*	2
	3.4. GLCM-correlation,	*CORR*	2
	3.5. GLCM-homogeneity	*HOMOGEN*	2
	3.6. GLCM-contrast	*CONTRAST*	2
	3.7. GLCM-dissimilarity	*DISSIMIL*	2
	3.8. GLCM-energy	*ENERGY*	2
4. Spectral-A	4.1. Integrated Energy for *k*-domain ~0.01–0.21 corresponds to wavelength ~10–2000 m (dependent on mode)	E_{ISP}	2
	4.2. Energy integrated with noise deduction (No Noise)	E_{NN}	2
	4.3. Rosenthal parameter (Energy integrated, scaled by $1/k$)	E_R	2
	4.4. Integrated Energy of a spectrum for wavelength 0–30 m	E^{30}	2
	4.5. Integrated Energy of a spectrum for wavelength 30–100 m	E^{100}	2
	4.6. Integrated Energy of a spectrum for wavelength 100–400 m	E^{400}	2
	4.7. Integrated Energy of a spectrum for wavelength 400–600 m	E^{600}	2
	4.8. Integrated Energy of a spectrum for wavelength 6000–2000 m	E^{2000}	2
	4.9. Integrated Energy of a spectrum for wavelength > 2000 m	$E^{>2000}$	2
	4.10. Spectrum Noise inside of *cut-off* domain of the spectrum	N_{in}^{S}	2
	4.11. Spectrum Noise outside of *cut-off* domain of the spectrum	N_{in}^{S}	2
	4.12. Energy max in the spectrum	E_{MAX}	2
	4.13. Longuet-Higgins spectral width parameter	P_{LH}	2
	4.14. Goda peakedness parameter	P_G	2
	4.15. Convolution of energy k_y and k_x axes projection	*CONV*	3
	4.16. Relation between positive and negative spectrum projection's differences	*REL*	3
	4.17. Relation between integrated energy k_y and k_x projection	S_{yx}	3
5. Spectral-B	5.1–5.20. 20 parameters as product of normalized image spectrum and 20 orthonormal functions	S_1– S_{20}	1
	5.21. Cutoff by ACF (Auto-Correlation-Function)	λc	1

parameters are presented in the appendix: T_{m0} and T_{m1} in Fig.A5, $H_s^{swell-1}$ and H_s^{swel-2} in Fig.A6, and H_s^{wind} and T^{wind} in Fig.A7.

All validations are performed using MFWAM monthly data blocks. In the scatterplots, only active point pairs are presented, the detected "bad values" are flagged out. For a month, the number of active points for the partially integrated parameters, e.g. for $H_s^{swell-1}$ is lower (for $H_s^{swell-2}$ even lower) than for total integrated H_s due to a high percentage of $H_s^{swell} = 0$ in ground truth. Note that, as the primary objective of this work was improving the main sea state parameter H_s, more data were used especially for this parameter. Despite these differences, for all parameters, validations were performed with a data amount three times exceeding the required minimum of ca. 20,000 collocations.

The S1 WV function was designed for VV polarization. The HH polarization data of around 5% of all acquisitions, validated separately, resulted in an RMSE ca. 10% lower than for VV data.

Each S1 IW and EW are acquired either in HH or VV polarization, distributed evenly. The accuracy of the developed universal model function is ca. 10 cm higher for VV than for HH polarization data. Since this effect was observed on both S1 IW and S1 EW, this shows that obviously at coarse resolution more sea state information is lost in SAR imagery in HH polarization compared to VV.

For TS-X data, the distribution between HH and VV polarization was around 60/40%. Based on S1 IW and S1 EW results, the algorithm was

Table 3
Cross validation of total significant wave height H_s using different combinations of ground truth data sources for tuning and validation. RMSE in meters is given for wv1/wv2 imagettes.

Tuning data source (setup)	Validation data source		
	MFWAM	WW3	BUOYS
MFWAM	0.33 / 0.38	0.35 / 0.40	0.42 / 0.44
WW3	0.34 / 0.39	0.34 / 0.39	0.44 / 0.46
MFWAM&WW3	0.34 / 0.39		0.43 / 0.45

Table 4
CWAVE_EX RMSE for eight integrated sea state parameters compared to model hindcasts.

№	Parameter	Unit	Satellite-mode			
			S1 IW	S1 EW	S1 WV (wv1/wv2)	TS-X SM
1	H_s	m	0.57	0.61	0.34 / 0.38	0.36
2	T_{m0}	s	0.91	0.86	0.46 / 0.51	0.72
3	T_{m1}	s	0.97	0.84	0.51 / 0.56	0.59
4	T_{m2}	s	0.82	0.86	0.46 / 0.51	0.51
5	$H_s^{swell-1}$	m	0.68	0.63	0.42 / 0.47	0.33
6	H_s^{swel-2}	m	0.38	0.44	0.41 / 0.46	0.27
7	H_s^{wind}	m	0.77	0.66	0.43 / 0.46	0.37
8	T^{wind}	s	0.97	0.95	0.62 / 0.67	0.71

Table 5
CWAVE_EX RMSE distribution (noise) for different sea state conditions for H_s.

Sea state domain, H_s (m)	Satellite-mode							
	S1-IW		S1-EW		S1-WV averaged wv1-wv2		TS-X SM	
	Fraction (%)	RMSE (m)	Fraction (%)	RMSE (m)	Fraction (%)	RMSE (m)	Fraction (%)	RMSE (m)
0.0–1.5	28	0.42	10	0.60	11	0.38	60	0.32
1.5–3.0	36	0.44	54	0.42	63	0.29	34	0.38
3.0–6.0	29	0.72	32	0.82	24	0.44	6	0.55
6.0 <	7	1.31	4	1.48	2	0.93	0	–
Total	100	0.57	100	0.61	100	0.35	100	0.36

Table 6
Total accuracy reached for H_s using the machine learning approach SVM relative to model and NDBC.

Data source	Total RMSE, m		Total BIAS, m		Non-valid images, %	
	wv1	wv2	wv1	wv2	wv1	wv2
MFWAM	0.245	0.273	−0.01	−0.01	1.9	1.4
NDBC	0.415	0.412	−0.10	−0.09	1.8	2.3

tuned for VV and HH polarization independently, resulting in an accuracy that was nearly identical for both polarizations.

The RMSE distribution for H_s and for different sea state domains is presented in Table 5. As can be seen, the distribution of wave heights is different for each mode. This is a consequence of differences in the area of the acquisitions for the modes. So, despite the huge database, not a single case with waves exceeding 6 m was collected for TS-X, acquired exclusively in coastal areas.

In the results, the accuracy decreases generally with increasing H_s values. This is connected to both: the accuracy of the SAR method and to the increased uncertainty in the ground truth data. However, the scatter index SI for each domain remains at around 10–15% if the local RMSE is connected with the mean value for this sea state domain. For low sea state, some difficulties can be seen for $H_s < 1.5$ m, where the accuracy does not follow the explanation above. This effect is related to the specialty of SAR imaging of the sea surface: the short and small waves cannot be imaged individually, but are only visible as image noise. Although it is possible to derive their characteristics from the SAR image noise, the accuracy is slightly lower than for more developed sea states.

5. Further algorithm improvement by machine learning

The following chapter deals with the further improvement of the sea state parameter estimation by applying machine learning techniques to S1 WV data.

5.1. Machine learning: Function and features

The fact that allowed machine learning to be used in the framework of this study was the access to the entire S1 WV Level-l archive. For earlier works with linear regression only individual S1 overflights in the order of one hundred thousand samples could be stored and considered, while for machine learning, the whole S1 WV archive with millions of samples was processed (see Table A3 data sets N° 4–8).

After a series of experiments with different machine learning approaches and also using knowledge gained in works performed on the same data (Stopa and Mouche, 2017; Quach et al., 2020), the support vector machine (SVM) technique was applied using the regression ν-SVR (Support Vector Regression nu-SVR, Chang and Lin, 2002) with a radial basis function as kernel-type. This function was chosen as the most suitable in terms of H_s-RMSE for the features and ground truth used, as a result from the testing of several different machine learning approaches with different kernel functions (e.g. epsilon-SVR, linear and sigmoid kernel types) applied to a relatively small number of collocations (100,000 samples). The SVR hyperparameters, which control the learning process, are: the cost regularization parameter *C,* the exponent degree in the radial basis function *g* (gamma) and the parameter ν (nu) which determines the proportion of the number of support vectors to keep in the solution with respect to the total number of samples in the dataset ($0 \leq \nu \leq 1$, $\nu = 0.5$ was implemented). For the tolerance of the termination criterion, which defines how close the compared models should be reached by the next iteration, the optimal value of $e = 0.01$ was found experimentally, then used for all trainings.

For practical applications, the high-performance ThunderSVM (TSVM) library was applied that runs an order of magnitude faster than the standard LibSVM (Wen et al., 2018). Using TSVM allows training of large datasets with millions of samples.

For training, all first-order normalized SAR features are applied. The feature normalization procedure was tested again using different methods. The optimal normalization method was found to be the same

Table 8
Accuracy distribution (uncertainty) for H_s using the machine learning approach SVM for different sea state domains and in total compared with NDBC buoy measurements averaged for S1 wv1 and wv2.

H_s domain (m)	N collocations	H_s fraction (%)	RMSE (m)	BIAS (m)
0.0–1.5	2539	29.0	0.39	−0.25
1.5–3.0	4431	50.6	0.36	−0.09
3.0–6.0	1637	18.7	0.48	0.07
6.0 <	151	1.7	0.93	0.57
Total	8757	100	0.41	−0.09

Table 7
Accuracy distribution (uncertainty) for H_s using the machine learning approach SVM for different sea state domains and in total compared with hindcast model data.

H_s domain (m)	H_s fraction (%)	SVM RMSE (m)		SVM BIAS (m)		SVM improvements against CWAVE_EX, averaged wv1/wv2	
		wv1	wv2	wv1	wv2	RMSE (cm)	|BIAS| (cm)
0.0–1.5	11	0.280	0.342	−0.07	−0.11	6.9	10.5
1.5–3.0	62	0.196	0.227	−0.01	−0.01	7.9	5.0
3.0–6.0	24	0.304	0.332	−0.01	0.02	12.2	1.3
6.0 <	2	0.519	0.558	0.03	0.02	39.2	16.1
Total	100	**0.245**	**0.273**	**−0.01**	**−0.01**	**9.1**	**4.5**

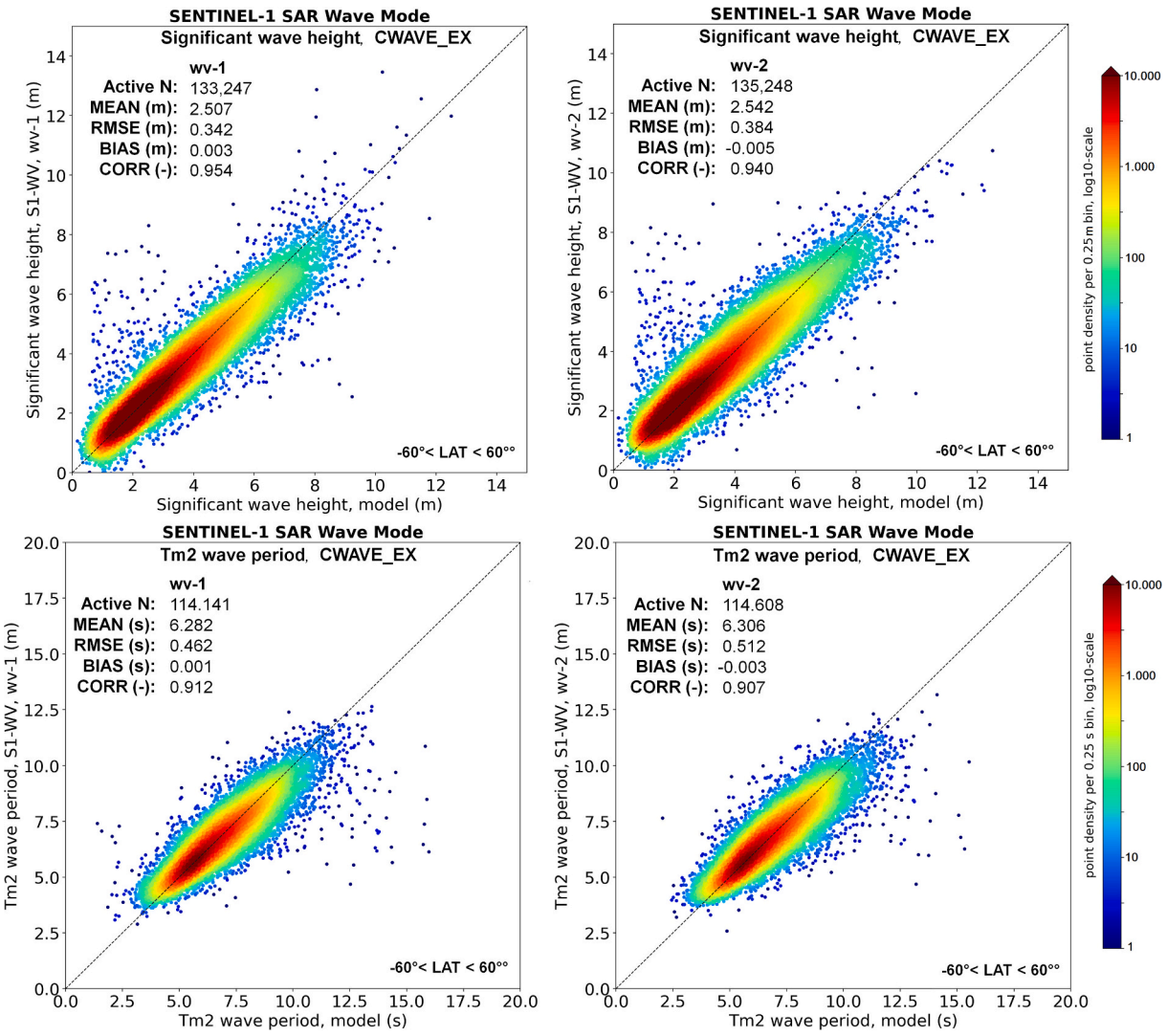


Fig. 2. CWAVE_EX results for S1 WV for H_s (first row) and T_{m2} (second row). The results are plotted separately for wv1 (left) and wv2 (right) imagettes.

as for linear regression (using MEAN and STD, see section 4.1 "Applying the CWAVE linear regression"). After a series of experiments, in order to improve the resulting accuracy and to shorten the training time, some additional features were included in the list of primary SAR features used for the linear regression approach before:

- first-guess H_s from linear regression solution CWAVE_EX.
- precise incidence angle (degree with accuracy of third decimal place).
- flag identifying the satellite (S1-A or S1—B).

5.2. Hyperparameters tuning

As a first step, the hyperparameters were tuned for a series of values of the cost parameter C in the range [25; 200] and gamma g in range [0.0010; 0.0100]. The hyperparameters' trainings were all carried out using the same data: 500,000 samples for each wv1 and wv2 for training (Table A3: **S1-WV_SVM_HP)** and ca. 300,000 samples for each wv1 and wv2 for test validation (Table A3: **S1-WV_SVM_TEST).**

The training consists of 20 independent trainings with different combinations of C and g (C = {25; 35; 55; 100; 200}, g = {0.0010; 0.0050; 0.0075; 0.0100}) where the resulting H_s-RMSE and training time are the output parameters. Fig. 4 displays the hyperparameters training results; the optimum combination found is C = 55 and g = 0.0075 (marked with a red circle). This combination was applied further for the final trainings of the model functions.

5.3. Training SVM models

For the final training of the SVM models (independently for wv1 and wv2), the optimal hyperparameters are applied. The training data consists of around one million samples for each of wv1 and wv2 (all data from June 2016 until December 2017, dataset **S1-WV_SVM_TR**), with all of the remaining data from Dec. 2018 until Feb. 2021 (ca. 4.3 million in total for both wv1 and wv2) used for the final validation (dataset **S1-WV_SVM_VAL**). The trainings were made in parallel for the wv1 and wv2 model functions SVM_SWH_wv1 and SWM_SWH_wv2 and took ca. two months. It is important to note that increasing the numbers of data samples increases the training time exponentially.

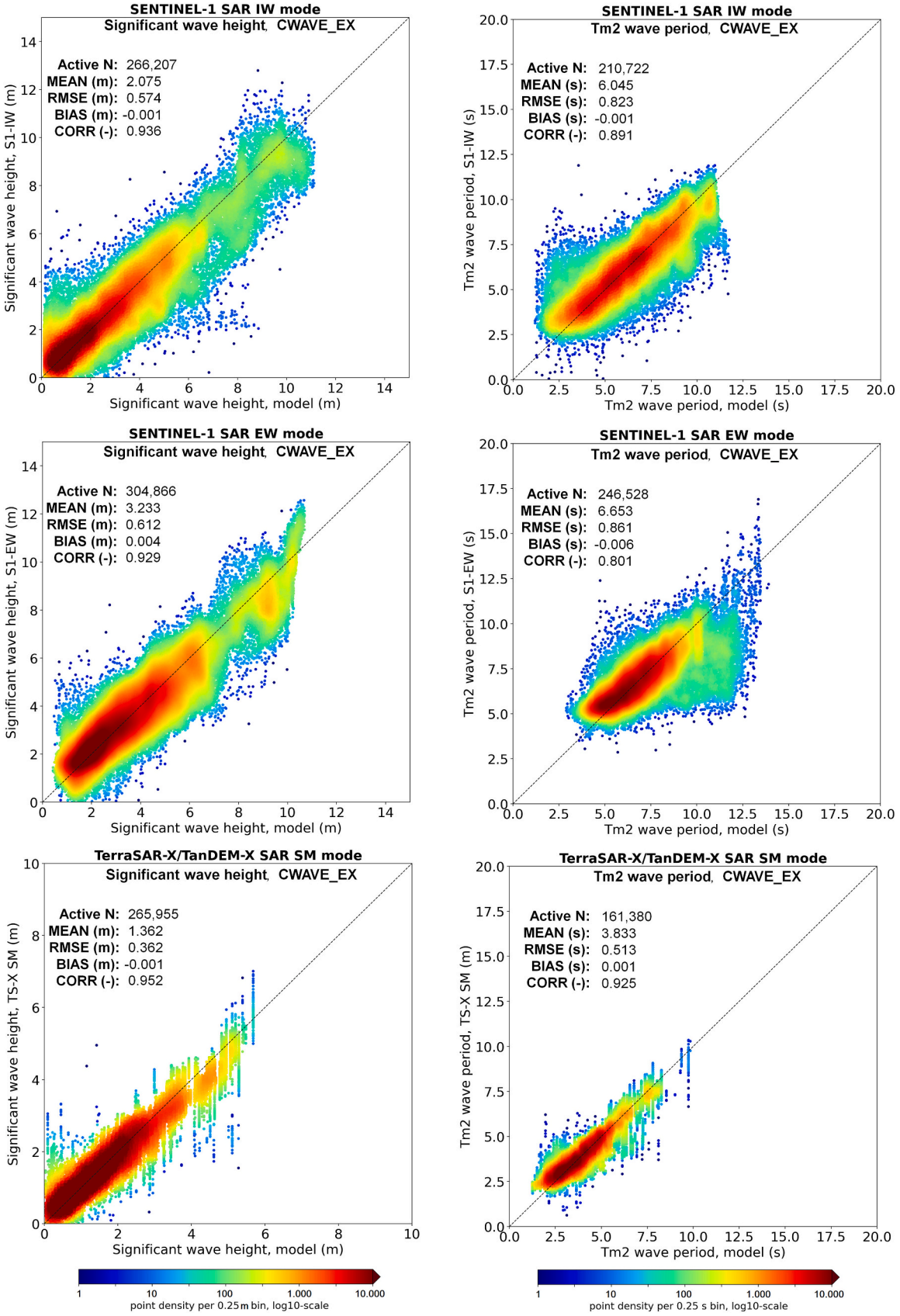


Fig. 3. CWAVE_EX results: H_s (left column) and T_{m2} (right column) for S1 IW (first row), S1 EW (second row) and TS-X (last row). In the 30 km × 50 km TS-X 1.5 km raster results the local distribution by comparison to coarse model grid points is visible.

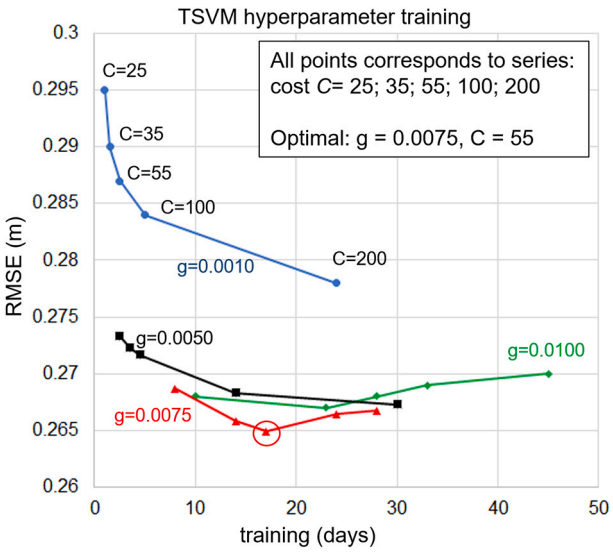


Fig. 4. Training of hyperparameters C (cost) and g (gamma) for nu-SVR regression with radial basis function kernel-type. The results present 20 independent trainings with different combinations of C and g; outputs are 20 pairs of H_s-RMSE and training time (axes). Each colour-line corresponds to one value of g and a series of C. The optimum in terms of accuracy and training time for the combination $\{g = 0.0075; C = 55\}$ is marked by a red circle and used for the final trainings. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.4. Sentinel-1 wave mode SVM results and comparison to CWAVE_EX

Applying the trained SVM model functions SVM_SWH_wv1 and SVM_SWH_wv2 to the validation data set **S1-WV_SVM_VAL** shows a total RMSE improvement of around 10 cm in comparison to the linear regression CWAVE_EX applied to the same data. The H_s accuracies reached RMSE = 0.245 m for wv1 imagettes and RMSE = 0.273 m for wv2 imagettes, when validated with model data (CMEMS, 2022). The improvements bring the accuracy of the results to the magnitude of the ground truth data uncertainty (around 25 cm, see ground truth comparisons in Section 2.5. and Fig.A1).

Fig. 5 shows the results for the H_s validation for the RR-2019 dataset (Table A3: **S1-WV_RR**) using the linear approach CWAVE_EX (left) and using the SVM technique (right). It is important to note that a series of outliers with overestimation of H_s for the model domain $0 < H_s < 3$ m was found to be a consequence of sea ice occurrence in the Antarctic region at latitude$<-55°$. In the following validations for the complete S1 WV archive, after excluding this area, such outliers are not present.

5.5. Processing the S1 WV archive in the scope of the ESA climate change initiative CCI

Until this section, only S1 WV orbits collocating with buoys were considered, which are around 5% of all acquired data. Using the SSP, the complete archive of S1 WV from December 2014 until February 2021 was processed. All processed S1 WV data including derived sea state parameters, imagette information (geo-location, time, ID, orbit number, etc.) and corresponding accuracy are stored both as ascii and in *netcdf* format for convenient use. The data is made available to the public within the scope of ESA's climate change initiative (CCI). Fig. 6 shows an example of Sentinel-1 Wave Mode WV archive processing for one day (right half) and one month (left half).

The validation of the whole archive (dataset **S1-WV_ARCHIVE**, Table A3) using hindcast model data (CMEMS, 2022) for latitudes of $-55° <$ latitude$<60°$ (avoiding ice coverage) resulted in an RMSE of 0.245/0.273 m for wv1/wv2 imagettes, respectively. The comparisons using NDBC buoys for 2017–2020 with almost 9000 collocations resulted in an RMSE = 0.41 m averaged over wv1 and wv2.

Table 6 presents the accuracy reached for H_s using SVM compared with model data and buoy measurements. Table 7 displays the details of the RMSE and BIAS distribution for different sea state domains and a comparison (improvements) to the accuracy of CWAVE_EX linear regression results.

Fig. 7 shows the SVM results for H_s validation for all S1 WV data since

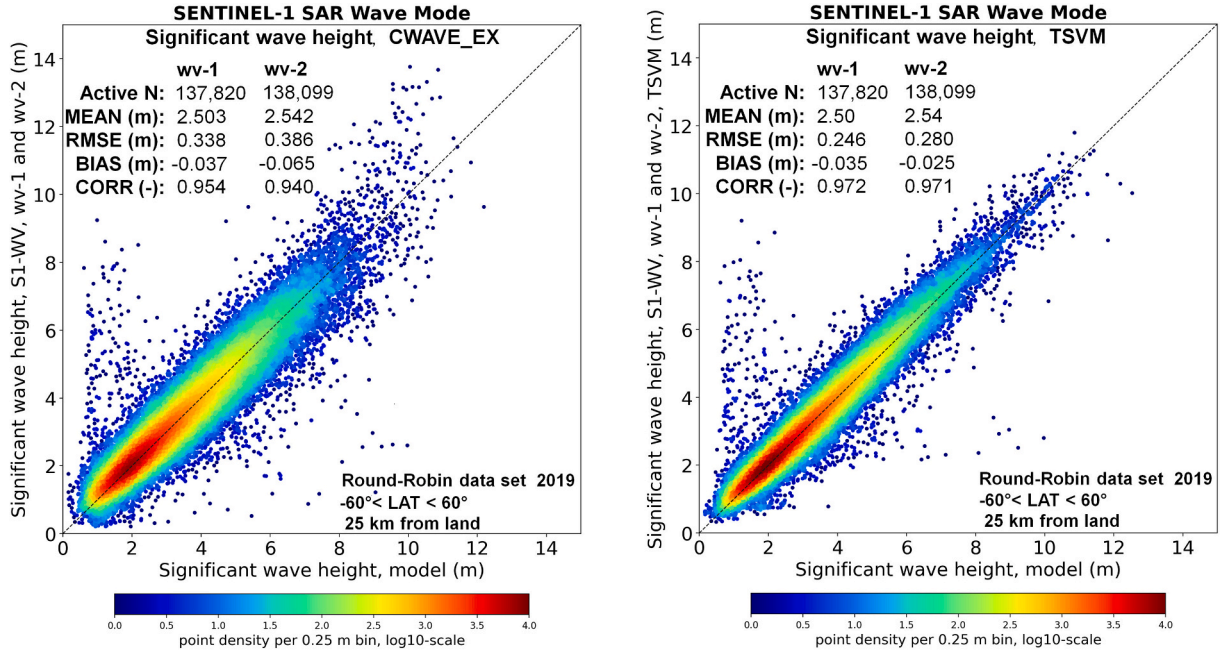


Fig. 5. H_s validation for the Round-Robin dataset (S1 WV buoy collocated orbits for 2019) with around 275,000 collocations using the linear approach CWAVE_EX (left) and the improved method using machine learning SVM (right).

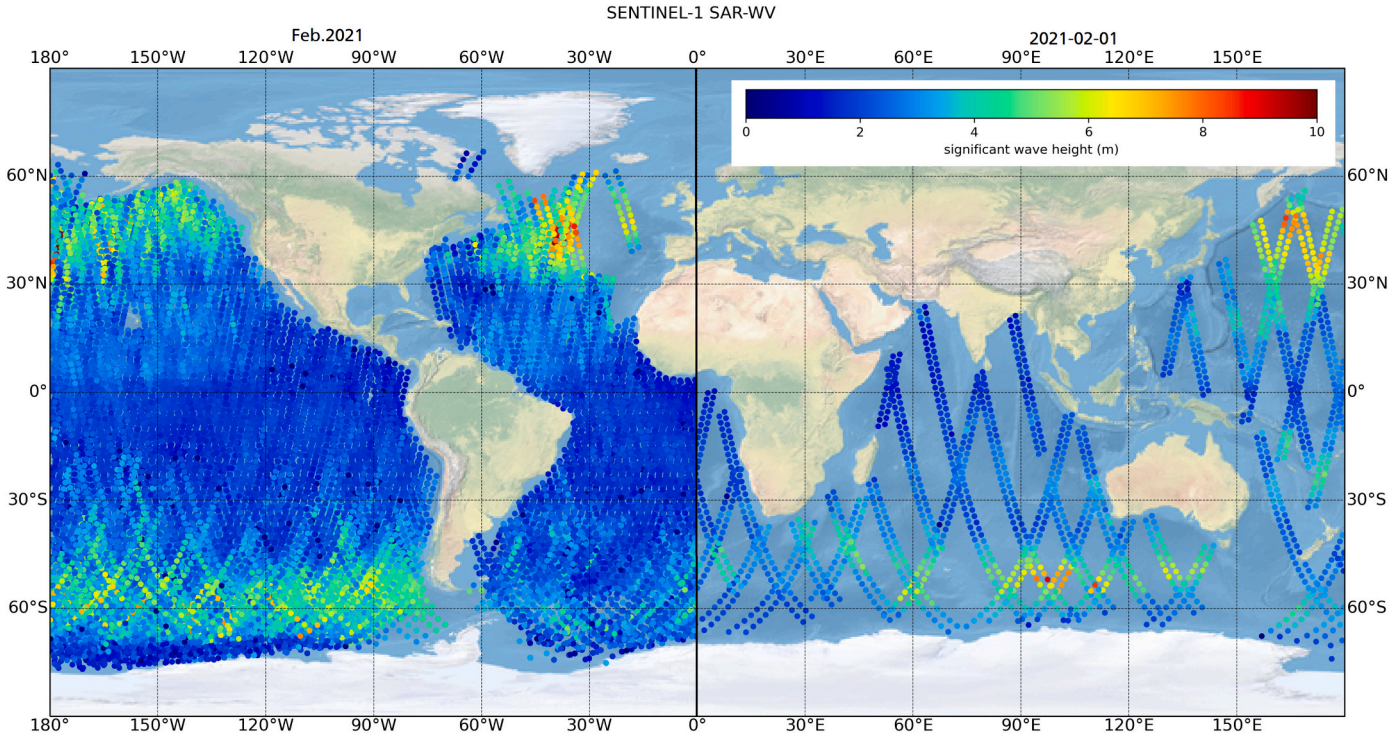


Fig. 6. Example of Sentinel-1 Wave Mode WV archive processing. On the right half of the globe only one-day of acquisitions is displayed, on the left half all data acquired during February 2021.

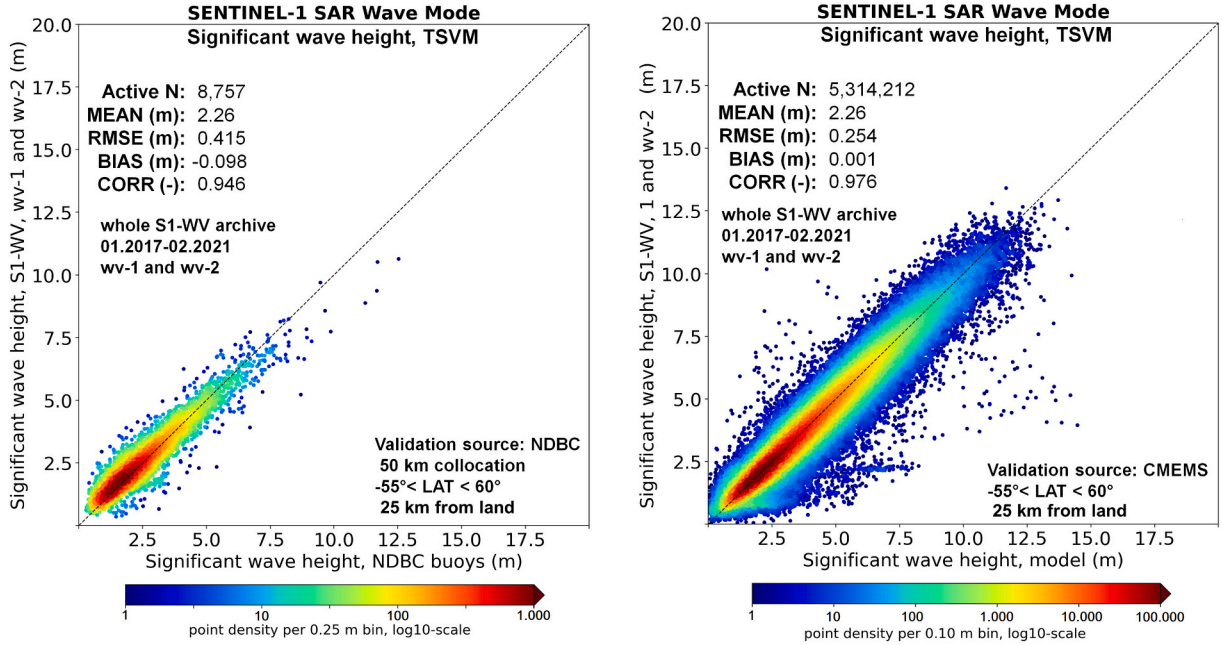


Fig. 7. TSVM H_s validation for all acquired S1 WV archive Jan. 2017 – Feb. 2021. NDBC buoys with almost 9000 collocations (left) and for MFWAM model (CMEMS, 2022) and with around 5 million collocations (right). The statistics is presented averaged for wv1 and wv2 imagettes.

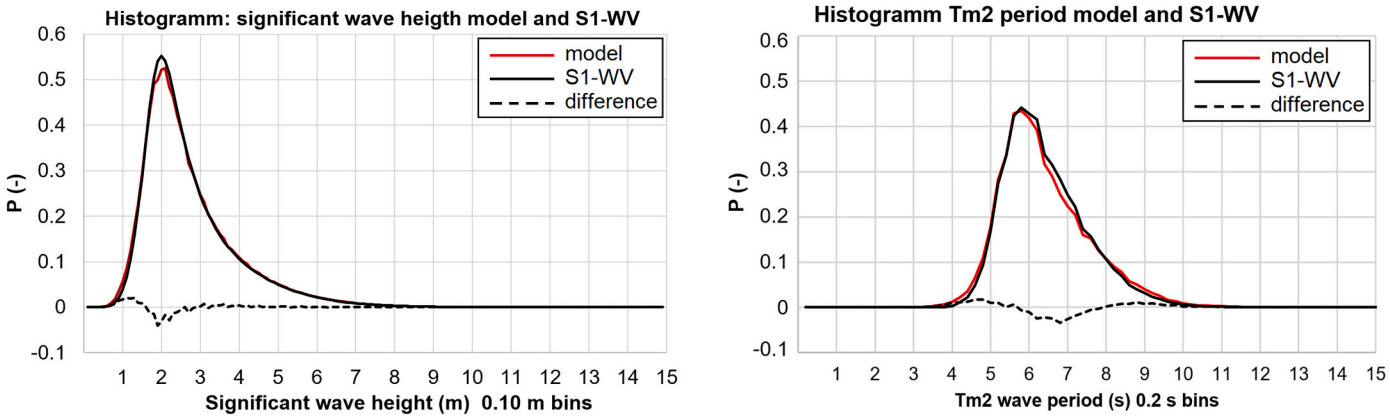


Fig. 8. Histograms H_s (TSVM) and T_{m2} (CWAVE_EX).

2017 for NDBC buoys (left) and for model data (right). Fig. 8 shows the histogram for H_s (SVM) and T_{m2} (CWAVE_EX) distributions. It is important to note that, despite minor differences, there are no domains where the SAR-estimated H_s was smoothed in PDF in comparison to the ground truth. This is an important result showing that new SAR methods work for all sea state domains quite accurately, especially for H_s under 1 m. Together with satellite altimeter, which is "affected by inadequate sampling of the signal at low sea states and is less accurate and is noisier particularly below 0.75 m" (Dodet et al., 2020), this allows to the study of the global sea state on a new qualitative level.

The resulting RMSE distribution (Table 7) across H_s domain SWH_RMSE (m) can be linearly approximated as:

$$\text{SWH\_RMSE} = 0.20 + 0.036\, H_s \tag{8}$$

As can be seen in Table 7, compared to the linear regression approach (Table 5), a significant improvement of around 40 cm in RMSE was reached for high sea state ($H_s > 6$ m).

When studying statistical means, in comparison to model data (CMEMS, 2022), BIAS = 0.001 m is quite accurate. However, in comparison to NDBC, the BIAS shows an overestimation of around 10 cm (BIAS = Ground_Truth – Estimated). For each domain, BIAS has the same sign by comparisons with buoys and model, however, for buoys the values are higher. The resulting BIAS = -0.09 m is dominated by the overestimation for middle sea state under 3 m (see Table 8).

6. Summary

This chapter summarizes the main findings in terms of accuracy improvements and method implementation.

6.1. General conclusions in terms of methods

The series of additional processing operations and new features significantly increases the accuracy of the H_s estimation compared to the original CWAVE. It also allows estimating wave periods, partial swell, and windsea parameters even for low-resolution SAR modes. The studies with a series of modifications conducted in this work show that the cumulative improvement in H_s-RMSE reached around 15 cm in comparison to CWAVE applied to the same S1 WV data and ground truth (Stopa and Mouche, 2017, RMSE~50 cm). This cumulative effect is composed as follows, sorted according to processing/implementation:

- Pre-filtering of artefacts (ships, slicks, etc) for all satellites and modes, improving the RMSE significantly in coastal areas by around 30 cm and slightly in the open ocean by around 1 cm (see Section 3.3. "Subscenes preparation") for all considered sensors.
- A cutting of the S1 WV imagette into subscenes and excluding non-valid subscenes improves the results by around 4 cm. Adding >9 subscenes (3 × 3, range×flight) does not result in further RMSE improvements (see Section 3.3. "Subscenes preparation").
- The resampling and smoothing of subscenes for low-resolution modes S1 IW and EW improves the RMSE by around 20 cm (see Section 3.3. "Subscenes preparation").
- The additional features introduced in the study result in an additional RMSE improvement of around 5 cm (Section 3.5. "Image spectrum and integration" and 3.6. "New features").
- Using additional inverse features within linear regression results in an additional improvement of around 3 cm (Section 4.1. "Applying the CWAVE linear regression function, extended algorithm CWAVE_EX").
- The SAR feature normalization results in an additional improvement of around 3 cm (Section 4.1. "Applying CWAVE linear regression function, extended algorithm CWAVE_EX")

It also became obvious that although the linear regression approach reaches a high accuracy with an RMSE~35 cm for H_s, further improvements by introducing additional features into CWAVE can hardly be achieved. The next step for accuracy improvement is machine learning, whereby the solution of the linear regression can be used as a first guess and applied as an additional feature for training. This approach slightly improves the results and shortens the training time.

In this work, the support vector machine (SVM) technique improves the H_s results by a further 10 cm compared to the CWAVE_EX (around 10%). A limited test for the Tm_2 period, performed with 100,000 wv1 samples, resulted in a comparable improvement (RMSE = 0.39 s against current value of 0.46 s).

6.2. Accuracy for different satellites and modes

The studies and comparisons conducted for S1 and TS-X show that the method tuned for different satellites and modes reaches an accuracy which depends more on the SAR image pixel resolution than on the radar band (C-band for S1 and X-band for TS-X) or satellite altitude (ca. 700 km for S1 and ca. 500 km for TS-X): An RMSE of around 0.35 m was

reached for both S1 WV and TS-X SM which have similar pixel spacing (2–4 m dependent on product and incidence angle), and an RMSE of around 0.60 m was achieved for the lower resolution S1 IW and S1 EW (pixel spacing of 10 m and 40 m, respectively). A small validation case study conducted for TS-X ScanSAR (SC, pixel spacing 10 m), using 150 TS-X SC images, also showed an RMSE of ca. 0.60 m.

6.3. Linear regression and machine learning capabilities

The advantage of the empirical approaches based on linear regression model (LRM) in comparison to machine learning model (MLM) is that an analytical solution exists. The function coefficients can be tuned comparatively quickly, extensive machine learning training is not necessary. Although the linear solution is inferior in accuracy to that obtained by adding machine learning, this solution is already stable with around 1/10 samples needed for machine learning in case of a distribution of the values in the dataset (see sea state fractions in Table 7) near to the normal distribution. In the scope of this study, it was found that for a stable solution valid for worldwide applications, the linear regression approach CWAVE_EX needs around 70,000 random samples for tuning, if all 131 features are used, achieving an RMSE of 0.34 m. Only when using this number of samples, additional data included into the tuning does not affect the accuracy by validation. For the machine learning SVM approach with the more complex model, this amount needs to be larger with at least 600,000 samples reaching an RMSE of 0.24 cm.

However, the developed MLM is many orders of magnitude larger than LRM (list of coefficients). Its practical application is outperformed by a linear application in terms of parsing speed of the model, which is important for near real-time services. In addition to the training time (can take months), it takes longer to read and apply the MLM to the estimated SAR features. This point is important, as a migration of the sea state processing for direct installation on a satellite for on-board-processing has been developed (Wiehle et al., 2022). In this case, no huge SAR data will be transferred from a satellite to earth, where the processing has to be done, but only processed sea state parameters in NRT. This technology will significantly simplify the data transfer and reduces the time gap between acquisition and receiving the processed sea state products at e.g. a ship bridge.

CRediT authorship contribution statement

Andrey Pleskachevsky: Conceptualization, Methodology, Investigation, Writing – original draft. **Björn Tings:** Methodology, Software, Data curation, Writing – review & editing. **Stefan Wiehle:** Methodology, Software, Writing – review & editing. **James Imber:** Methodology, Writing – review & editing. **Sven Jacobsen:** Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data are available (ESA)

Acknowledgements

Parts of this work, namely the development of a postprocessor enabling SVM-training for improved estimation of wave height from preprocessed S1 WV features and the processing of S1 WV archive, were supported by the European Space Agency (ESA) within the CCI+ Sea State project (ESA, 2022a, CCI-SeaState). The authors are grateful to the team of the DLR ground station Neustrelitz for active cooperation and organization of the NRT service chain and providing processed data to the users. Special thanks go to Andreas Jahnke (DLR, EOC) for providing the Sentinel-1 WV data archive. The authors are also grateful to Susanne Lehner for organizing the research group and providing first scientific steps and Wolfgang Rosenthal (RIP 2016) for qualified advice concerning sea state, ship navigation and suggestions concerning algorithm developments.

Appendix A. Appendix

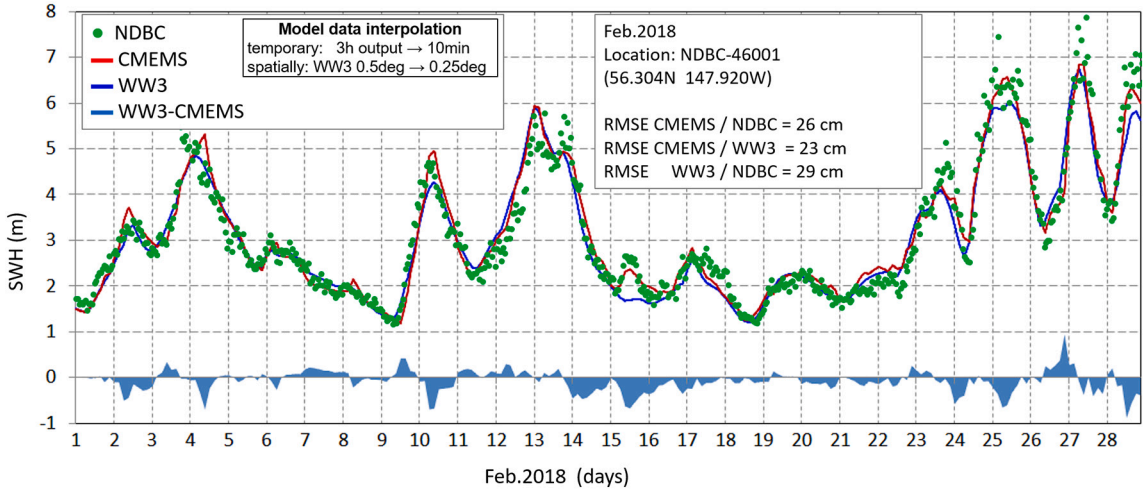


Fig. A1. Example comparison for H_s conducted for NDBC-46001 (56.304 N 147.920 W), MFWAM, and WW3 models for February 2018.

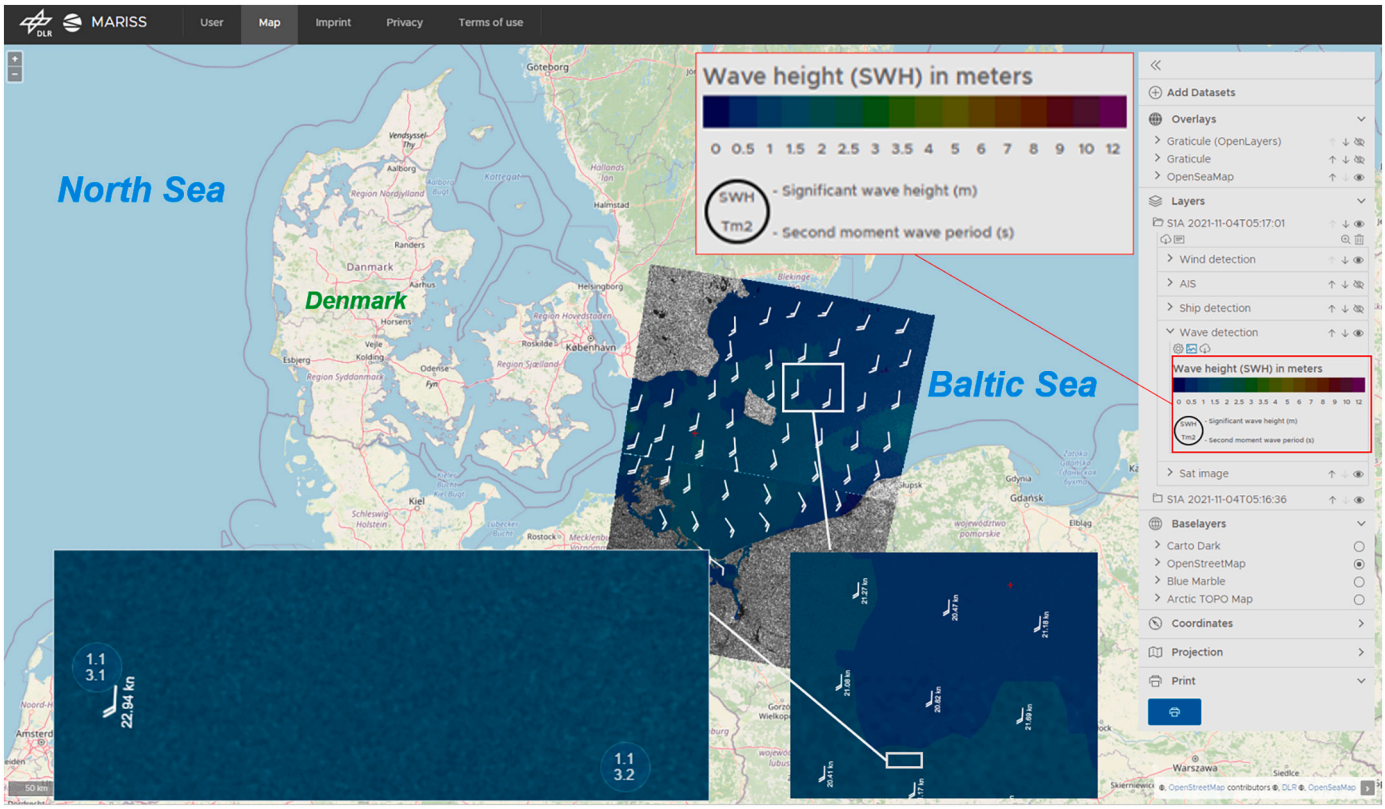


Fig. A2. Screenshot of the demonstrator for NRT services at Ground Station Neustrelitz. The demonstrator runs daily for Sentinel-1 IW in Southern North Sea and Western Baltic Sea. The actual processing raster is 3 km, the wave-detection layer shows wave height (colored) and period (in circles: H_s above, Tm_2 below). Data for all eight sea state parameters can be downloaded as Google Earth kmz file. The wind-detection layer shows the wind speeds estimated from the SAR image.

Table A1
SAR feature's criteria for validity of subscenes.

Feature	Satellite-mode			
	S1 IW	S1 EW	S1 WV	TS-X SM
MI	>0.100000	>0.100000	>0.100000	>0.100000
E_{ISP}	> 0.000500	> 0.000500	> 0.000500	> 0.000500
E^{75}	> 0.000005	> 0.000005	> 0.000005	> 0.000005
E^{390}	> 0.000100	> 0.000100	> 0.000100	> 0.000100
E^{600}	> 0.000500	> 0.000500	> 0.000500	> 0.000500
nv	> 1	> 1	> 1	> 1
E_{MAX}	<500	<500	<1000	<1000
CONTRAST	> 0.000010	> 0.000010	> 0.000010	> 0.000010
E_R	<3000	<3000	<5000	<850
LANDMASK	Land pixels<20%	Land pixels<20%	Land pixels<20%	Land pixels<20%

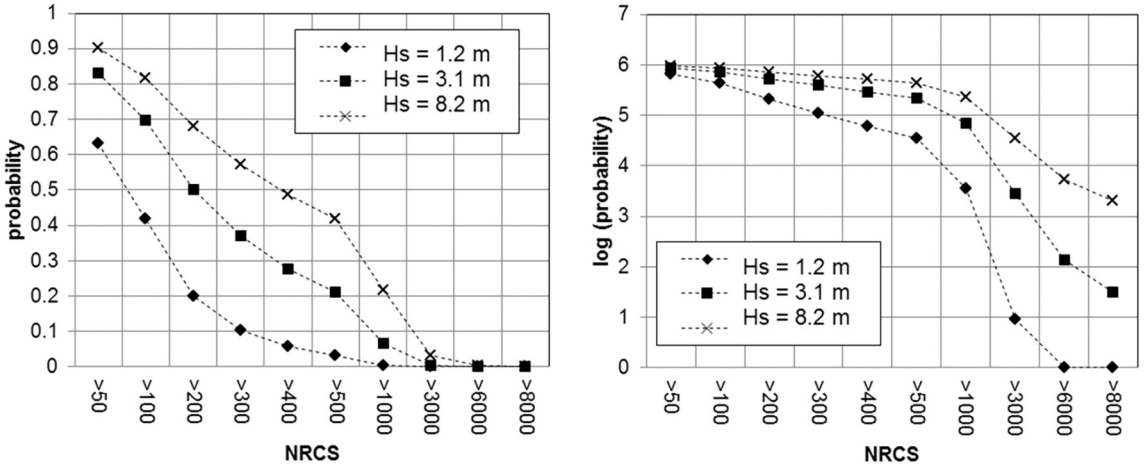


Fig. A3. An example of the probability distribution (Complementary Cumulative Distribution Function CCDF) of NRCS (S1 WV wv1) in subscenes for three typical cases: low sea state $H_s = 1.2$ m, medium sea state $H_s = 3.1$ m, and high sea state $H_s = 8.2$ m (left), and the same on a logarithmic scale (right).

Table A2
Weighting factor G for CCDF function with 21 bins.

NRCS value (bin number i)	Weighting factor G_i
50(1)	50
100(2), 200(3), 300(4), 400(5)	100
500(6), 1000(7), 1500(8), 2000(9), 2500(10), 3000(11)	500
4000(12), 5000(13), 6000(14), 7000(15), 8000(16), 9000(17), 10,000(18)	1000
12,000(19), 15,000(20)	2000
20,000(21)	5000

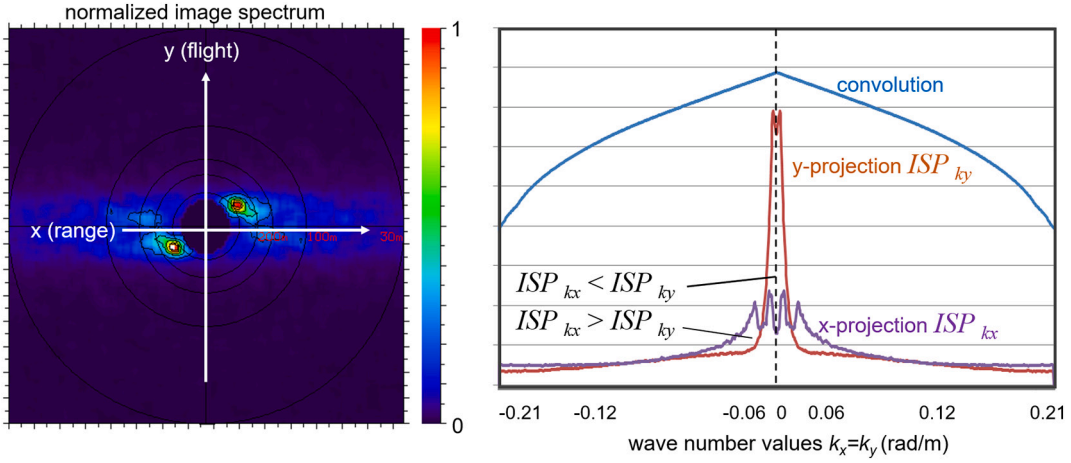


Fig. A4. An example of 1-D accumulative projections of a 2-D image spectrum onto k_y and k_x axes. Due to the *cutoff* effect, the peak in k_y-protection is stronger.

Table A3
Data processed for tuning, training and validation of S1 WV.

N°	Data set name	N° ID scenes	ground truth	N° of imagettes			Note
	Data set description			wv1	wv2	Total	
1	**S1-WV_LR_TUN**	2409	MFWAM, WW3	109,693	112,930	222,623	linear CWAVE_EX tuning
	NDBC collocated orbits 2015–2017						
2	**S1-WV_LR_VAL**	2489	MFWAM, WW3	137,780	136,431	222,623	linear CWAVE_EX validation
	NDBC collocated orbits 2018						
3	**S1-WV_RR** (RoundRobin)	2430	MFWAM, WW3	135,967	138,099	274,066	CWAVE_EX TSVM validation
	NDBC collocated orbits 2019		NDBC	1041	963	2004	CWAVE_EX TSVM validation
4	**S1-WV_SVM_HP**	23,780	MFWAM	500,000	500,000	1000,000	SVM hyper parameters training
	whole archive 06.2016–06.2017						
5	**S1-WV_SVM_TEST**	18,890	MFWAM	300,000	300,000	600,000	SVM test
	whole archive 06.2017–12.2017						
6	**S1-WV_SVM_TR**	37,690	MFWAM	1,080,496	1,065,639	2,146,135	SVM training
	whole archive 06.2016–12.2017						
7	**S1-WV_SVM_VAL**	76,320	MFWAM	2,133,062	2,135,326	4,268,388	SVM validation
	whole archive 12.2018–02.2021						
8	**S1-WV_ARCHIVE**	161,923	–	6,7 Mio	6,7 Mio	13,4 Mio	sea state parameters archive
	Whole S1 WV archive 12.2014–02.2021		MFWAM05.2016–0.22021	2,524,250	2,789,962	5,314,212	archive validation MFWAM
			NDBC 01.2017–12.2020	4302	4406	8757	archive validation NDBC

Table A4
Data processed for tuning and validation of S1 IW, EW and TS-X SM, ground truth MFWAM.

Used data	No. of products	No. of collocated subscenes total	No. of collocated subscenes tuning Data set name	No. of collocated subscenes validation Data set name
S1 IW	1762	517,289	300,000 **S1-IW_TUN**	217,289 **S1-IW_VAL**
S1 EW	2093	1,162,492	800,000 **S1-EW_TUN**	362,492 **S1-EW_VAL**
TS-X SM	2047	216,938 (HH pol.)	120,000 (HH pol.) **TS-X_HH_TUN**	96,938 (HH pol.) **TS-X_HH_VAL**
		138,885(VV pol.)	120,000 (VV pol.) **TS-X_VV_TUN**	18,885 (VV pol.) **TS-X_VV_VAL**

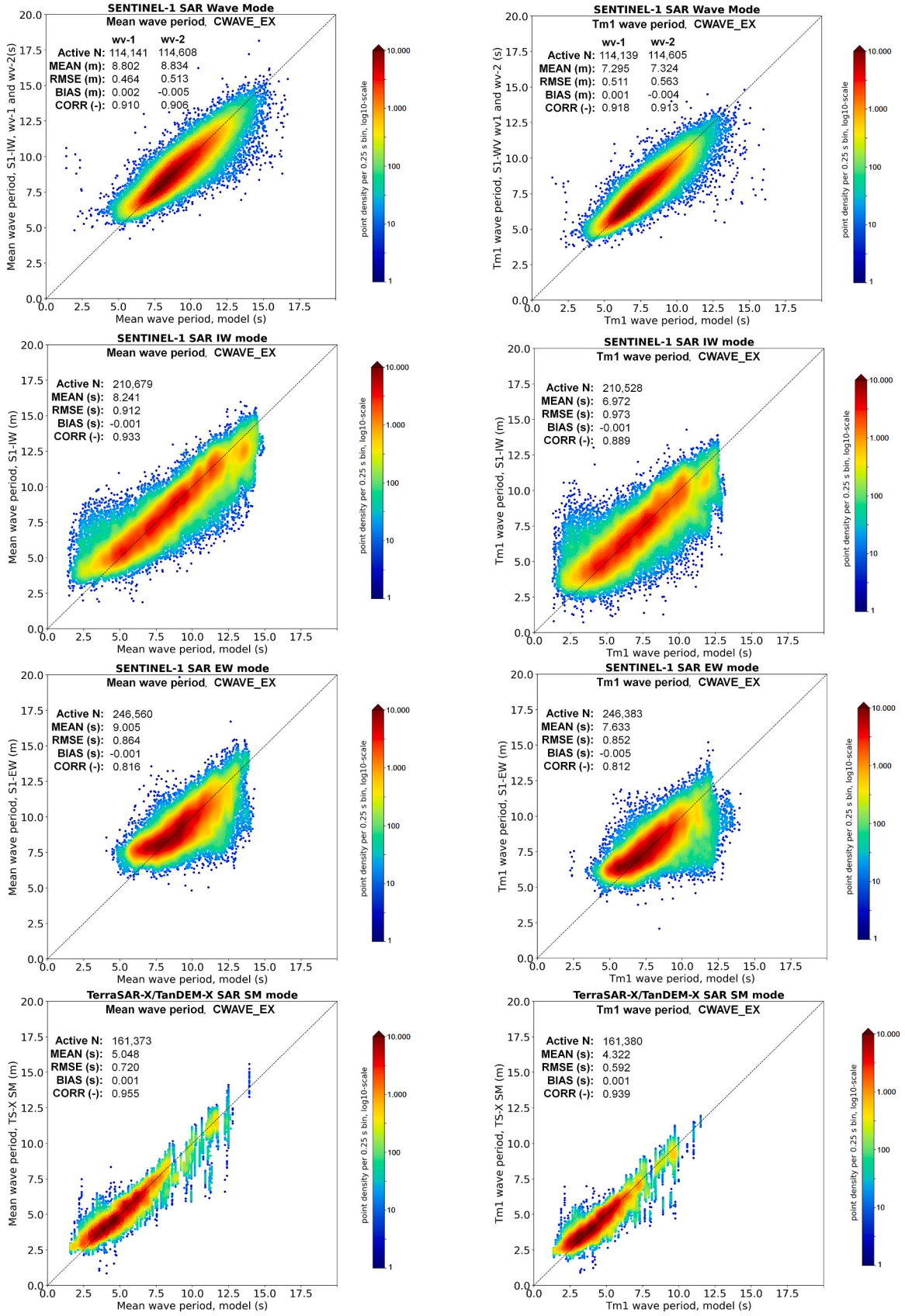


Fig. A5. CWAVE_EX: scatterplots for periods T_{m0} (left column) and T_{m1} (right column) for S1 WV including both wv1 and wv2 (first row), S1 IW (second row), S1 EW (third row) and TS-X (last row). In the 30 km × 50 km TS-X 1.5 km raster results the local distribution by comparison to coarse model grid points is visible.

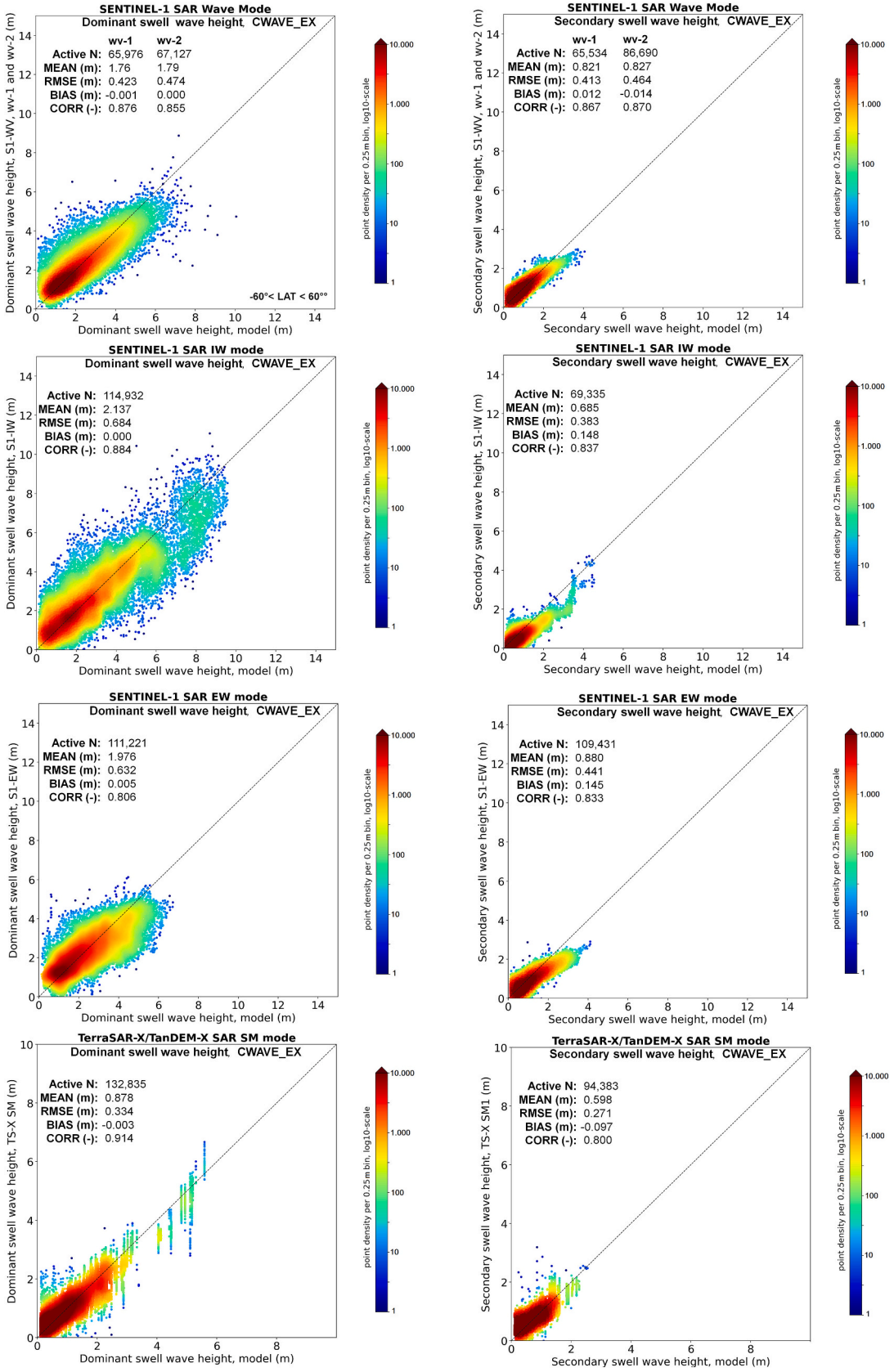


Fig. A6. CWAVE_EX: scatterplots for partially integrated dominant swell $H_s^{\text{swell-1}}$ (left column) and $H_s^{\text{swel-2}}$ (right column) for S1 WV including both wv1 and wv2 (first row), S1 IW (second row), S1 EW (third row) and TS-X (last row). In the 30 km × 50 km TS-X 1.5 km raster results the local distribution by comparison to coarse model grid points is visible.

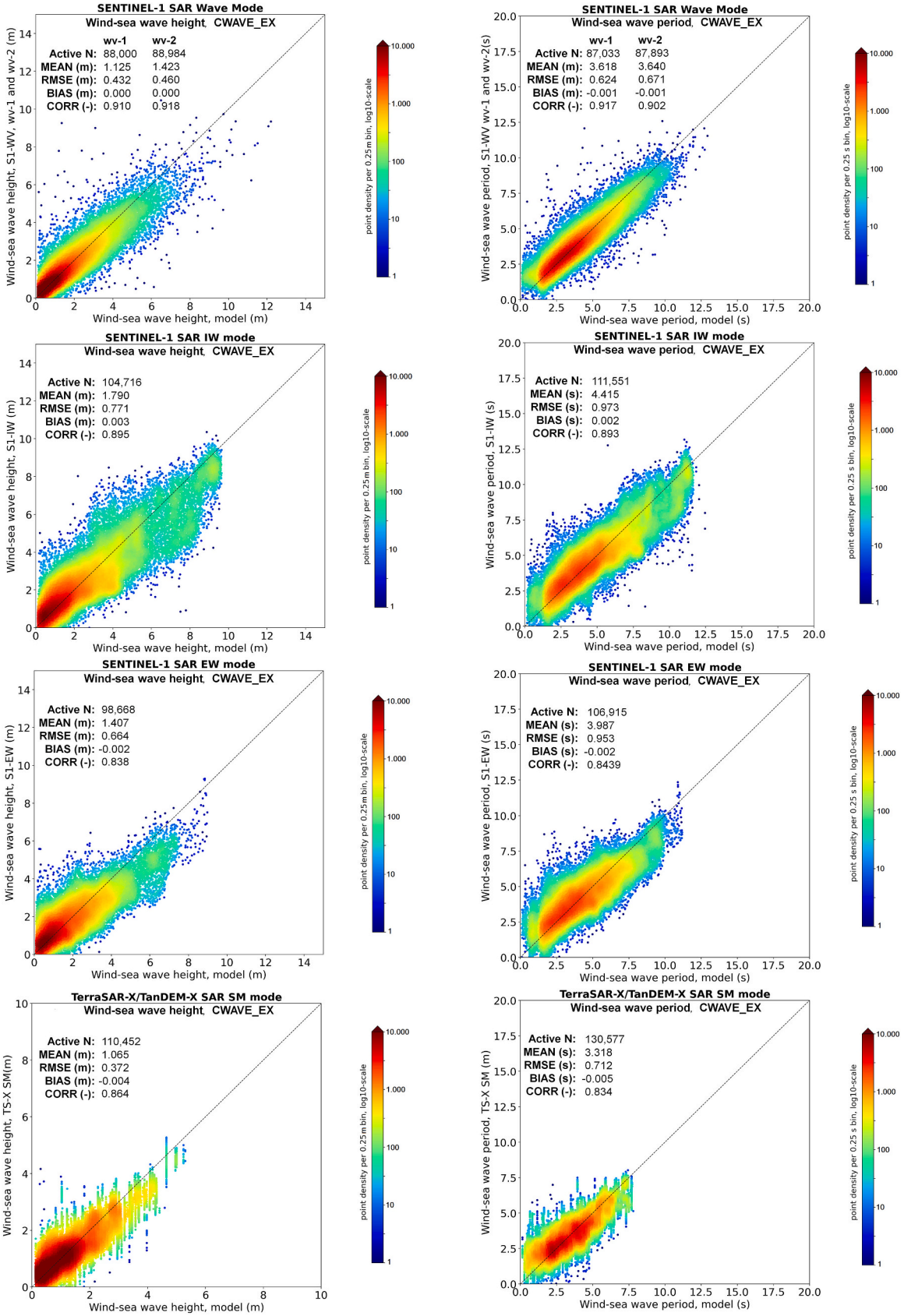


Fig. A7. CWAVE_EX: scatterplots for partially integrated wind-sea wave height H_s^{wind} (left column) and wind-sea wave period T^{wind} (right column) for S1 WV including both wv1 and wv2 (first row), S1 IW (second row), S1 EW (third row) and TS-X (last row). In the 30 km × 50 km TS-X 1.5 km raster results the local distribution by comparison to coarse model grid points is visible.

References

Abdalla, S., Bidlot, J., Janssen, P., 2006. Global Validation and Assimilation of Envisat ASAR Wave Mode Spectra. European Space Agency, (Special Publication) ESA SP.

Abdalla, S., Bidlot, J., Janssen, P., 2010. Envisat asar wave mode spectra global validation and assimilation. In: Proc. 'SeaSAR 2010', Frascati, Italy, 25–29 January 2010. ESA SP-679, April 2010.

Alpers, W., Rufenach, C., 1979. The effect of orbital motion on synthetic aperture radar imagery of ocean waves. IEEE Trans. Antennas Propag. 27, 685–690. https://doi.org/10.1109/TAP.1979.1142163.

Ardhuin, F., Magne, R., Filipot, J.-F., Van der Westhyusen, A., Roland, A., Quefeulou, P., Lefèvre, J.M., Aouf, L., Babanin, A., Collard, 2010. Semi empirical dissipation source functions for wind-wave models: Part I, definition and calibration and validation at global scales. JPO 40, 1917–1941.

Ardhuin, F., Stopa, J., Chapron, B., Collard, F., Smith, M., Thomson, J., Doble, M., Blomquist, B., Persson, O., Collins, C., Wadhams, P., 2016. Measuring ocean waves in sea ice using SAR imagery: a quasi-deterministic approach evaluated with Sentinel-1 and in situ data. Remote Sens. Environ. 189, 211–222.

Bruck, Miguel, 2015. Sea State Measurements Using TerraSAR-X/TanDEM-X Data. PhD Theses. University of Kiel, Kiel.

Chang, C.-C., Lin, C.-J., 2002. Training ν-support vector regression: theory and algorithms. Neural Comput. 14 (8), 1959–1977.

CMEMS, 2022. Copernicus Marine Environment Monitoring Service. https://marine.copernicus.eu/.

Copernicus Resources, 2022. https://resources.marine.copernicus.eu/product-detail/GLOBAL_ANALYSIS_FORECAST_WAV_001_027/DOCUMENTATION.

Dodet, G., Piolle, J.-F., Quilfen, Y., Abdalla, S., Accensi, M., Ardhuin, F., Ash, E., Bidlo, J.-R., Gommenginger, C., Marechal, G., Passaro, M., Quartly, G., Stopa, J., Timmermans, B., Young, J., Cipollini, P., Donlon, C., 2020. The sea state CCI dataset v1: towards a sea state climate data record based on satellite observations. Earth Syst. Sci. Data 12, 1929–1951.

ECCC, 2022. http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/waves-vagues/index-eng.htm.

EMODNET, 2022. http://www.emodnet-physics.eu/Map/#.

ESA, 2022a. CCI-SeaState. Sea State Project. http://cci.esa.int/seastate (accessed on 13 April 2020).

ESA, 2022b. Sentinel-1. https://sentinel.esa.int/web/sentinel/missions/sentinel-1.

ESA, 2022c. Manual. https://sentinels.copernicus.eu/web/sentinel/radiometric-calibration-of-level-1-products.

ESA, 2022d. S1-EW. https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/extra-wide-swath.

ESA, 2022e. S1-IW. https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/interferometric-wide-swath.

ESA, 2022f. S1-WV. https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/wave.

ESA, 2022g. TS-X. https://earth.esa.int/web/eoportal/satellite-missions/t/terrasar-x.

Frost, A., Wiehle, S., Singha, S., Krause, D., 2018. Sea ice motion tracking from near real time sar data acquired during antarctic circumnavigation expedition. In: Proc. IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 2338–2341.

Goda, Y., 1970. Numerical experiments on wave statistics with spectral simulation. Rep. Port Harbour Res. Inst. 9, 3–75.

Hasselmann, S., Brüning, C., Hasselmann, K., Heimbach, P., 1996. An improved algorithm for the retrieval of ocean wave spectra from SAR image spectra. J. Geophys. Res. 101-7, 16615–16629.

Hersbach, H., 2008. CMOD5.N: A C-band Geophysical Model Function for Equivalent Neutral Wind.

Holt, B., 2004. SAR imaging of the ocean surface. In: Jackson, C.R., Apel, J.R. (Eds.), Synthetic Aperture Radar (SAR) Marine User's Manual. NOAA NESDIS Office of Research and Applications, Washington DC, pp. 25–79.

Kerbaol, V., Chapron, B., 1998. Analysis of ERS-1/2 synthetic aperture wave mode imagettes. JRS 103-4, 7833–7846.

Lehner, S., Pleskachevsky, A., Bruck, M., 2012. High resolution satellite measurements of coastal wind field and sea state. Int. J. Remote Sens. 33-23, 7337–7360.

Lehner, S., Pleskachevsky, A., Velotto, D., Jacobsen, S., 2013. Meteo-marine parameters and their variability observed by high resolution satellite radar images. J. Oceanogr. 26–2, 80–91. https://doi.org/10.5670/oceanog.2013.36.

Li, X.-M., Huang, B., 2020. A global sea state dataset from spaceborne synthetic aperture radar wave mode data. Sci. Data 7, 261. https://doi.org/10.1038/s41597-020-00601-3.

Li, X.-M., Lehner, S., 2014. Algorithm for sea surface wind retrieval from TerraSAR-X and TanDEM-X data. IEEE Trans. Geosci. Remote Sens. 52 (5), 2928–2939. https://doi.org/10.1109/TGRS.2013.2267780.

Longuet-Higgins, M.S., 1984. Statistical properties of wave groups in a random sea-state. Phil. Trans. R. Soc. Lond. A 249, 321–387.

NDBC, 2022. https://www.ndbc.noaa.gov/.

NOAA, 2022. https://polar.ncep.noaa.gov/waves/.

NOAA Manual, 2022. https://polar.ncep.noaa.gov/waves/wavewatch/.

Pleskachevsky, A., Rosenthal, W., Lehner, S., 2016. Meteo-marine parameters for highly variable environment in coastal regions from satellite radar images. JPRS 119, 464–484.

Pleskachevsky, A., Jacobsen, S., Tings, B., Schwarz, E., 2019. Estimation of sea state from Sentinel-1 synthetic aperture radar imagery for maritime situation awareness. IJRS 40-11, 4104–4142.

Pleskachevsky, A., Jacobsen, S., Tings, B., Schwarz, E., Krause, D., Voinov, S., 2022. Multiparametric Sea State Fields from synthetic aperture radar. EUSAR-2022. In: Proc. 14th European Conference on Synthetic Aperture Radar, 23–27.07.2022, Leipzig, Germany, p. 6.

Quach, B., Glaser, Y., Stopa, J., Mouche, A., Sadowski, P., 2020. Deep learning for predicting significant wave height from synthetic aperture radar. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2020.3003839.

Ressel, R., Frost, A., Lehner, S., 2015. A neural network-based classification for sea ice types on X-band SAR images. IEEE J. Select. Top. Appl. Earth Observ. Rem. Sens. 8-7, 3672–3680.

Rikka, S., Pleskachevsky, A., Jacobsen, S., Alari, V., 2018. Meteo-marine parameters from Sentinel-1 SAR imagery: towards near real-time services for the Baltic Sea. Remote Sens. 10 (5), 757, 17p.

Schulz-Stellenfleth, J., König, Th., Lehner, S., 2007. An empirical approach for the retrieval of integral ocean wave parameters from synthetic aperture radar data. JRL 112, 1–14.

Schwarz, E., Krause, D., Berg, M., Daedelow, H., Maas, H., 2015. Near real time applications for maritime situational awareness. In: Proc. 36th International Symposium on Remote Sensing of Environment, 11–15 May 2015, Berlin, Germany Remote Sensing and Spatial Information Sciences, XL-7/W3, p. 6.

Shao, W., Zhang, Z., Li, X., Li, H., 2016. Ocean wave parameters retrieval from Sentinel-1 SAR imagery. Remote Sens. 8 (9), 707.

Singha, S., Vespe, M., Trieschmann, O., 2013. Automatic synthetic aperture radar based oil spill detection and performance estimation via a semi-automatic operational service benchmark. Mar. Pollut. Bull. 15-73 (1), 199–209.

Stopa, J., Mouche, A., 2017. Significant wave heights from Sentinel-1 SAR: validation and applications. JGR 122, 1827–1848.

Sun, M., Yang, Y., Yin, X., Du, J., 2018. Data assimilation of ocean surface waves using Sentinel-1 SAR during typhoon Malakas. Int. J. Appl. Earth Obs. Geoinf. 70, 35–42.

Tings, B., 2021. Non-linear modeling of detectability of ship wake components in dependency to influencing parameters using Spaceborne X-band SAR. Remote Sens. 13-2, 165.

Tings, B., Bentes, C., Velotto, D., Voinov, S., 2018a. Modelling ship detectability depending on TerraSAR-X-derived meteo-ocean parameters. CEAS Space J. 11-1, 81–94.

Tings, B., Pleskachevsky, A., Velotto, D., Jacobsen, S., 2018b. Extension of ship wake detectability model for non-linear influences of parameters using satellite based X-band synthetic aperture radar. Remote Sens. 11-5 (563), 20.

Tings, B., Pleskachevsky, A., Velotto, D., Jacobsen, S., 2019. Extension of ship wake detectability model for non-linear influences of parameters using satellite based X-band synthetic aperture radar. Remote Sens. 11-5, 20p.

Voinov, S., Schwarz, E., Krause, D., Tings, B., 2020. Earth Observation Maritime Surveillance System. Proc. GeoForum MV 2020, 31 August - 1 September 2020, Rostock. ISBN 978–3–95545-337-4. GITO Verlag, pp. 73–77.

Wen, Z., Shi, J., Li, Q., He, B., Chen, J., 2018. ThunderSVM: a fast SVM library on GPUs and CPUs. J. Mach. Learn. Res. 19, 1–16.

Wiehle, S., Breit, H., Günzel, D., Mandapati, S., Balss, U., 2022. Synthetic aperture radar image formation and processing on an MPSoC. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/TGRS.2022.3167724.