Product Specification Document (PSD)

version 2.0, 14 October 2021

Contents

Author	Approved	Signature	Date
JF Piolle, G Dodet, B Tings, A Pleskachevsy	Ellis Ash Fabrice Ardhuin		14/10/21
ESA Approval	Digitally signed by Paolo Cipollini Date: 2021.11.25 18:37:58 +01'00'		

Issue	Date	Comments
0.5	14 Dec 2018	Full draft for internal review
1.0	05 Feb 2019	First version for review by ESA
1.1	23 Apr 2019	Various minor updates
1.2	17 Jun 2019	Updates prior to release of v1 dataset
2.0	14 October 2021	Updates for the release of v2 dataset

1. Introduction

This document presents the Product Specification Document (PSD) for **Sea_State_cci**, deliverable 2.0 of the project. This version is prepared in order to define the version 2 products that have been released in August 2021, updating and extending products from the version 1 released in 2019. It also provides a basis for the Product User Guide. An updated version of this document, describing further versions of products, will be produced towards the end of the project.

The remainder of this Product Specification Document contains a general description of the version 2 products and provides the specific variable content of the datasets. Annexes describe the detailed netCDF format, differences with v1 data, and processing details.

2. General Products Description

2.1 Products overview

The 2nd version (version 2) of CCI Sea State products provides a complete revised processing and methodology compared to the version 1 which was mainly inherited from the GlobWave project. It extends and improves the altimeter significant wave height (SWH) products which were a post-processing over existing L2 altimeter agency products with a complete new retracking of the considered missions for this version 2, and fully revised filtering, corrections and uncertainty evaluations. It also adds SAR Wave Mode measurements to the dataset.

Three kinds of products are delivered:

- L2P : Along-track products from altimetry and SAR missions separated per satellite and pass, including all measurements with flags, corrections and extra parameters from other sources. These are expert products with rich content and no data loss.
- L3 : Edited merged daily products retaining all valid and good quality measurements from all altimeters over one day (one daily file), with simplified content (only a few key parameters). This is close to what is delivered in NRT by CMEMS project.
- L4 : Gridded products averaging valid and good measurements from all available altimeters over a fixed resolution grid (1°x1°) on a monthly basis. These products are meant for statistics, and visualization through the CCI toolbox.

2.2 Altimeter missions

For version 2, a full retracking of altimeter waveforms was performed with two different retrackers: WHALES by TUM for LRM altimeters, and LR-RMC by CLS/CNES for SAR altimeters. Because of the novelty of this approach and the required processing effort, this was only applied to a limited set of recent altimeters spanning from 2000 to 2020.

As a reminder, version 1 dataset included older missions spanning from 1991 to 2018. The list of altimetry missions included in version 2 dataset are:

Mission	Altimeter	Selected band	Source product	Temporal coverage
Envisat	RA-2	Ku	SGDR v3 [ESA]	14-May-2002 to 8-Apr-2012

Jason-1	POSEIDON-2	Ku	SGDR version E [Aviso]	15-Jan-2002 to 3-Mar-2012
Jason-2	POSEIDON-3	Ku	SGDR version D [Aviso]	4-Jul-2008 to 17-May-2017
Jason-3	POSEIDON-3B	Ku	SGDR version D [Aviso]	17-Feb-2016 to 1-Jun-2019
Cryosat-2	SIRAL	Ku	SIR LRM L1B version D [ESA]	16-Jul-2010 to 8-Jul-2020
Saral	AltiKa	Ka	SGDR version T [Aviso]	14-Mar-2013 to 11-Nov-2019

2.3 SAR missions

SAR Wave Mode data are a new addition to this CCI Sea State version 2 Dataset, with various complementary products providing significant wave height and directional parameters for recent missions embedding a SAR with wave mode.

Mission	SAR	Band	Source product	Temporal coverage
Envisat	ASAR	C	ESA WVI	10-Dec-2002 to 8 Apr-2012
Sentinel-1 A	C-Band SAR	C	S1A_WV_SLC	1-Dec-2014 to 23-Feb-2021
Sentinel-1 B	C-Band SAR	C	S1B_WV_SLC	15-Jun-2016 to 23-Feb-2021

2.4 Format

All products are in NetCDF4 Classic format. They comply to the CCI Data Standard 2.1 (See http://cci.esa.int/working-groups) and follow the CF 1.7 and ACDD conventions

The detailed format of each product is described in the following sections.

2.5 Data access, organisation and file naming

The data can be obtained from Ifremer FTP server: ftp://eftp.ifremer.fr. The login and password can be obtained upon filling the registration form at :

https://forms.ifremer.fr/lops-siam/access-to-esa-cci-sea-state-data/

The common directory structure is based on CCI recommendations and is arranged as follows:

/products/<release version>/data/<instrument type>/<type>/<mission>/<date>/

Where:

- **<cci_project>** : seastate
- **<release version>** is the dataset version (currently 2.0.6 for this dataset)
- **<instrument type>** is the type of remote sensing technique: **altimeter** or **sar**
- **<type>** : will be different for each ECV, but needs to be defined, and consistent within an ECV, here **l2** for along-track data, **l3** for edited merged products and **l4** for monthly averaged gridded products)
- **<mission>** : satellite mission (for L2P products only)
- **<date>** : <year as YYYY>/<day in the year as DDD>

The file nomenclature is based on form 2 of CCI recommendations :

ESACCI-<CCI Project>-<Processing Level>-<Data Type>-<Product String>[-<Additional Segregator>]-<IndicativeDate>[<Indicative Time>]-fv<File version>.nc

Where:

- **<CCI Project>** : SEASTATE
- **<Processing Level>** : here L2P for along-track data, L3 for edited merged product and L4 for monthly averages
- **<Data Type>** : SWH for “Significant Wave Height”, ISSP for “Integrated Sea State Parameters”
- **<Product String>** : contains the name of the mission for L2P

- **<Additional Segregator>** : optionally the name of algorithm used
- **<Indicative Date>[<Indicative Time>] :** The identifying date for this data set. Format is YYYYMMDDTHHMMSS, where YYYY is the four digit year, MM is the two digit month from 01 to 12 and DD is the two digit day of the month from 01 to 31. The date used should best represent the observation date for the data set. For along-track data it will be the time of the first measurement in the file.
- **<File version>** : File version number in the form n{1,}[.n{1,}] (That is 1 or more digits followed by optional . and another 1 or more digits.)

Examples:

L2P altimeter product:

ESACCI-SEASTATE-L2P-SWH-Jason-2-20170130T145103-fv01.nc

L3 product:

ESACCI-SEASTATE-L3-SWH-MULTI_1D-20170130-fv01.nc

L4 product:

ESACCI-SEASTATE-L4-SWH-MULTI_1M-201701-fv01.nc

3. Altimeter Datasets

3.1 L2P

The altimeter L2P products are along-track files, usually corresponding to a satellite pass, processed from each mission data provider's L1B product (usually referred to as SGDR) containing the waveforms. A retracking algorithm is applied to retrieve significant wave height from the instrument waveforms.

Additional post-processing is performed, like quality control, adjustment of significant wave height to a common reference, uncertainty estimation, etc… and complementary variables are also computed or added from other model or satellite sources. The content is fully consistent and standardized for each mission included in this dataset.

For altimeters, only the Ku band measurements are considered, when available (the only current exception being Saral for which only Ka band is provided).

This section describes in detail the specific content of the version 2 L2P for Sea State CCI.

3.1.1 Content overview

The following table provides the list of variables included in the L2P product.

	coordinate variables
time	time (in seconds since 1985-01-01)
lat	latitude
lon	longitude
	instrumental variables for 1st band altimeter (usually Ku)
sigma0	Ku band backscatter coefficient, as calculated from the retracking
sigma0_rms	RMS of the Ku band backscatter coefficient, within 1Hz cells, of the 20 Hz measurements calculated from the retracking

sigma0_num_valid	number of valid points used to compute Ku band backscatter coefficient, within 1Hz cells, of the 20 Hz measurements calculated from the retracking
	environmental variables for 1st band altimeter (usually Ku)
swh	significant wave height, within 1Hz cells, averaged over the 20 Hz measurements calculated from the retracking
swh_adjusted	adjusted significant wave height. The adjustment is based on cross-mission intercalibration as described in section 4.1.2.5.
swh_denoised	EMD-filtered significant wave height, an adjusted and denoised significant wave height estimated by CCI Sea State project and based on Quilfen and Chapron, 2021. [Quilfen Y., Chapron B. (2020). On denoising satellite altimeter measurements for high-resolution geophysical signal analysis. *Advances in Space Research, 68.* https://doi.org/10.1016/j.asr.2020.01.005]
swh_denoised_uncertainty	In the EMD processing of an along-track data segment, the uncertainty attached to each denoised measurement in the segment characterizes its expected error, and is calculated as the local standard deviation of a set of twenty denoised Hs data segments obtained from twenty random realizations of the noisy Hs data segment (Quilfen and Chapron, 2021) [Quilfen Y., Chapron B. (2020). On denoising satellite altimeter measurements for high-resolution geophysical signal analysis. *Advances in Space Research, 68.* https://doi.org/10.1016/j.asr.2020.01.005]
swh_noise	High frequency noise on significant wave height, estimated by CCI Sea State project and based on Quilfen and Chapron, 2021. [Quilfen Y., Chapron B. (2020). On denoising satellite altimeter measurements for high-resolution geophysical signal analysis. *Advances in Space Research, 68.* https://doi.org/10.1016/j.asr.2020.01.005]
swh_quality	quality level if significant wave height measurements, as estimated using the editing criteria described in section 4.1.2.4.

swh_uncertainty	best estimate of the standard deviation of the significant wave height standard error, as described in section 4.1.2.6.
swh_rms	RMS of significant wave height within 1Hz cell of the unadjusted 20 Hz measurements calculated from the retracking
swh_num_valid	number of valid points used to compute the significant wave height, within 1Hz cells, from the 20 Hz measurements calculated from the retracking
swh_rejection_flag	flag specifying the editing criteria on which a measurement was rejected (meaning its quality level is not set to “good”).
	auxiliary variables
auxiliary measurements	
sla_unfiltered	sea level anomaly, as taken from CMEMS SEALEVEL_GLO_PHY_L3_REP_OBSERVATIONS_008_062 dataset
mdt	mean dynamic topography, as taken from CMEMS SEALEVEL_GLO_PHY_L3_REP_OBSERVATIONS_008_062 dataset
sea_ice_fraction	sea ice fraction, as taken from CCI Sea Ice products
total_column_liquid_water_content_rad	Total column cloud liquid water content from onboard radiometer, when available in source product.
topography	
bathymetry	ocean depth, taken from GEBCO version 20150318
distance_to_coast	distance to nearest coast, from NASA/GFSC distance to coastline
model auxiliary data	
wind_speed_model_u	U component of the model wind vector, taken from ERA5 reanalysis

wind_speed_model_v	V component of the model wind vector, taken from ERA5 reanalysis
sea_surface_temperature	sea surface temperature, taken from ERA5 reanalysis
surface_air_temperature	surface air temperature, taken from ERA5 reanalysis
surface_air_pressure	surface air pressure, taken from ERA5 reanalysis
total_column_liquid_water_content	Total column cloud liquid water content, taken from ERA5 reanalysis
swh_model	significant wave height, taken from ERA5 reanalysis

3.1.2 Processing details

3.1.2.1 Retracking

In CCI SeaState version 1 Dataset, significant wave height was taken from agencies GDR products. In this new version 2 of CCI Sea State Dataset, we have performed a complete retracking of the included missions from the wave forms available in input SGDR products (Level 1) provided by space agencies. The retracking calculates 20 Hz SWH measurements (40 Hz for Saral) from the waveforms. TUM's retracker WHALES was used for all non-SAR altimetry missions (Envisat, CryoSat-2, Saral, Jason-1, Jason-2, Jason-3). Version 3 of CCI Dataset will include SAR altimetry mission (Sentinel-3) retracked with CNES LR-RMC retracker.

The Low Resolution Mode (LRM) waveforms are characterised by a rising leading edge that becomes less steep as the SWH increases, and a slowly decreasing trailing edge. The standard retracking methods are still affected by a suboptimal distribution of the residuals in the fitting process, which results in high level of noise in the estimations. WHALES is designed as a unified way to solve these problems and is based on two principles:

1) The application of a weighted fitting solution, whose weights are adapted to the SWH in order to guarantee a more uniform distribution of the residuals during the iterative fitting. This guarantees significantly more precise estimations.

2) A subwaveform strategy to focus the retracking on the portion of the signal of interest, avoiding heterogeneous backscattering in the trailing edge (partially inherited from the ALES

retracker, Passaro et al., 2014[1]). This guarantees efficiency in the coastal zone and a better representation of the oceanic scales of variability.

3) The decorrelation between SWH and sea level estimation, which corrects for possible covariant errors and increases furthermore the precision.

Moreover, a revisiting of the look-up tables used to correct for the Gaussian approximation of the Point Target Response in the Brown model ensures that the accuracy in the estimation is tailored to the new retracking solution.

3.1.2.2 Correction of the 20 Hz measurements

Most altimeter retrackers assume that the emitted pulse can be approximated by a Gaussian (an exception is the numerical retracker developed by CLS). In practice the Gaussian shape is never achieved and thus there are errors in the inversion associated with this assumption. A common work-around is to apply the retracker to highly realistic simulations including the full Point Target Response (PTR) and note the differences between retrieved and simulated values. These can be used to generate a Look-up Table (LUT) giving the error i.e. the necessary correction to be applied to yield unbiased estimates. The Geophysical Data Records (GDRs) produced by the operational agencies contain the corrected value of SWH and the instrumental correction that was applied.

As the WHALES algorithm fits the Brown-Hayne model to the waveform shape, it was deemed appropriate to use the LUT already determined for the standard MLE-3/MLE-4 retracker applied to those waveform data. Note, the correction to be applied is not simply the value in the corrections field for that particular instant (as that will be using the MLE3/4 estimate to index the LUT); rather we use the same LUT (each one being specific to a particular altimeter) but determine the correction according to the WHALES estimate of SWH.

The LUTs pertinent to each instrument are not published, but may be derived by tabulating the values in the GDRs. These appear to be values of SWH only and unchanging throughout the mission; an exception is Envisat, where the value of the correction in the GDR varies with sigma0 (or with other co-varying parameters) and it also shows some change through the mission. As details of the correction model were not accessible, this has been approximated by a function of SWH only.

Caveats

I. For some GDRs the correction is only provided at 1 Hz and therefore tabulated against the 1 Hz values of SWH; for others it is available at 20Hz. Within the Sea

[1] Passaro M., Cipollini P., Vignudelli S., Quartly G., Snaith H.: ALES: A multi-mission subwaveform retracker for coastal and open ocean altimetry. Remote Sensing of Environment 145, 173-189, 10.1016/j.rse.2014.02.008, 2014

State CCI, it is being applied to the 20 Hz SWH estimates from WHALES and then these resultant values are used to produce 1 Hz data.

II. A retracker applied to noisy waveform data may sometimes lead to 20 Hz estimates below 0 m. These are valid and should be corrected as for slightly positive values before generating 1 Hz values. The correction value in the lowest SWH bin (0.00-0.10 m) should be used for negative SWH estimates

III. There is an expectation that the correction error would tend to a constant value for large values of SWH. This is clearly the case for Jason-1, Jason-2, Jason-3 and Envisat. The GDR-derived corrections for AltiKa and CryoSat-2 appear to be still increasing (becoming less negative) between 8 and 12 m. For implementation within the Sea State CCI, we set the corrections for each instrument to be constant above 13 m, using the last value in the LUT (13.00-13.10 m).

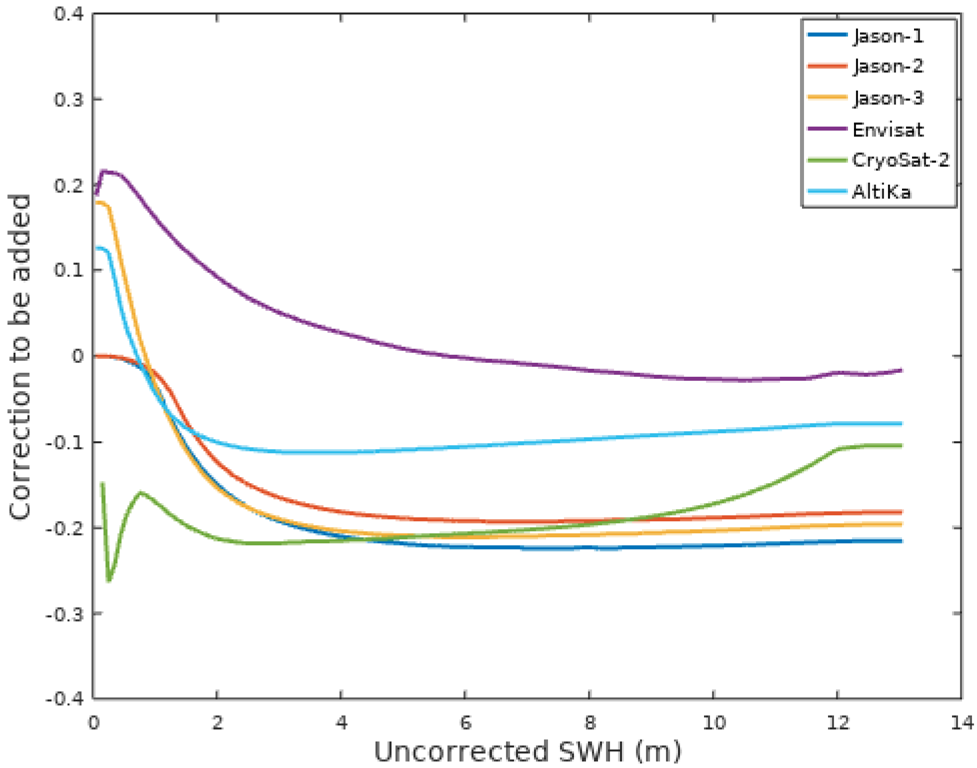

Figure 3.1: The PTR corrections for the 6 LRM altimeters used in the Sea State CCI v2.

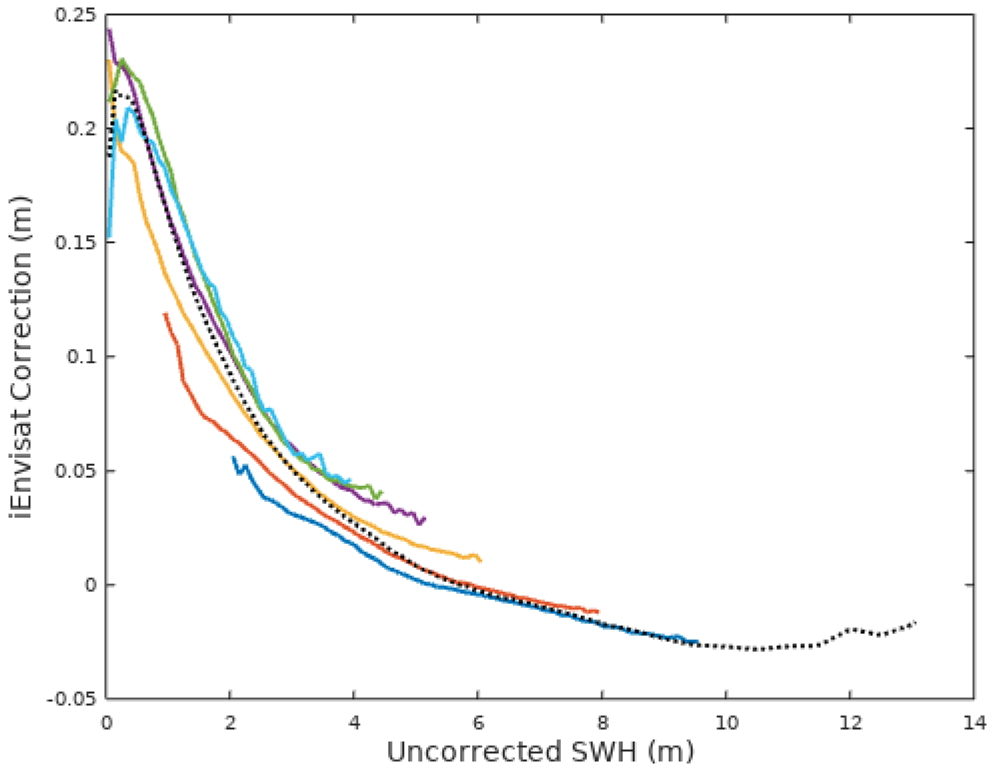

Figure 3.2: For Envisat, the PTR corrections on the GDRs also vary with sigma0 (coloured lines correspond to mean SWH correction at intervals of 1 dB. The black dotted line shows the overall mean correction, reflecting the change in typical sigma0 associated with increasing SWH.

3.1.2.3 Averaging of the 20 Hz measurements

The CCI Sea State Dataset v2 provides 1 Hz SWH measurements. These 1Hz measurements are calculated by averaging groups of 20 Hz (40 Hz for Saral):

- for Envisat, CryoSat-2, Saral, Jason-1, Jason-2, Jason-3 missions, the groups of 20 Hz measurements used to calculate the 1Hz values are exactly the same as in the source Agency's GDR & SGDR products. Both CCI and Agency files can be compared one to one, have the same number of measurements, and the same latitude, longitude, time for each 1 Hz measurement.
- for Sentinel3-A, the 20 Hz measurements were grouped by seconds to calculate the 1 Hz values: starting from the first measurement in the file, the next measurements separated by less than one second from the first one are grouped together, and so on. The nominal number of 20 Hz measurements per 1 Hz value is not constant and varies between 19 and 20. It can be less in case of missing data, end of orbit, etc...

The method used to average the 20 Hz measurements into 1 Hz values is the same for all altimeters. For each group of 20 Hz measurements:

1. 20 Hz measurements flagged as bad by the retracker are discarded. 20 Hz measurements are flagged as bad, for altimeters retracked with WHALES, when the fitting error is greater than 0.3.
2. 20 Hz measurements located on land are discarded. The land location is determined from the latitude & longitude provided for each measurement using OpenstreetMap coastline, with a precision of about 50m.
3. 20 Hz SWH measurements not in the -0.5 to 30 meter range are discarded
4. the remaining 20 Hz SWH outliers are discarded. The outlier detection scheme is based on the maximum absolute deviation for a 3-sigma criterion, meaning
   - only 20 Hz SWH values within [median(SWH) - 3 * MAD(SWH), median(SWH) 3 * MAD(SWH)] interval are kept
   - with: MAD(SWH) = 1.4286 * median(abs(SWH - median(SWH)))
5. the 1Hz swh value is computed as the median of the remaining 20Hz swh values that have not been discarded during the 4 previous steps. Moreover, the swh_num_valid parameter is computed as the number of remaining 20Hz records and the swh_rms parameter is computed as the root mean square of the deviation from the median, considering only valid 20Hz records

The same method is used to calculate 1 Hz sigma0 values from the 20 Hz sigma0 (except no test on the validity range of the sigma0 is performed).

This processing step result in the following coordinates and variables in the CCI Sea State Dataset v2:

time

```
double time(time) ;
        time:_FillValue = 1.e+20 ;
        time:long_name = "measurement time" ;
        time:standard_name = "time" ;
        time:units = "seconds since 1981-01-01T00:00:00+00:00" ;
        time:calendar = "proleptic_gregorian" ;
```

latitude [lat]

```
double lat(time) ;
        lat:_FillValue = 1.e+20 ;
        lat:long_name = "latitude" ;
        lat:standard_name = "latitude" ;
        lat:units = "degrees_north" ;
         lat:axis = "Y" ;
```

```
        lat:valid_range = -90, 90 ;
        lat:comment = "Positive latitude is North latitude, negative latitude is South latitude" ;
        lat:coverage_content_type = "coordinate" ;
```

longitude [lon]

```
double lon(time) ;
        lon:_FillValue = 1.e+20 ;
        lon:long_name = "longitude" ;
        lon:standard_name = "longitude" ;
        lon:units = "degrees_east" ;
        lon:axis = "X" ;
        lon:valid_range = -180, 180 ;
        lon:comment = "East longitude relative to Greenwich meridian" ;
        lon:coverage_content_type = "coordinate" ;
```

Significant wave height [swh]

Hs as retrieved from WHALES retracker [no calibration, no editing].

```
double swh(time) ;
        swh:_FillValue = 1.e+20 ;
        swh:comment = "All instrumental corrections included. As available from retracker and unedited." ;
        swh:least_significant_digit = 3LL ;
        swh:ancillary_variables = "swh_quality swh_rejection_flags" ;
        swh:coverage_content_type = "physicalMeasurement" ;
        swh:band = "Ku" ;
        swh:coordinates = "lat lon" ;
        swh:standard_name = "sea_surface_wave_significant_height" ;
        swh:authority = "CF 1.7" ;
        swh:long_name = "Ku band significant wave height" ;
        swh:units = "m" ;
```

significant wave height rms [swh_rms]

```
double swh_rms(time) ;
        swh_rms:_FillValue = 1.e+20 ;
        swh_rms:least_significant_digit = 3LL ;
        swh_rms:coverage_content_type = "auxiliaryInformation" ;
        swh_rms:band = "Ku" ;
        swh_rms:coordinates = "lat lon" ;
        swh_rms:authority = "CF 1.7" ;
```

```
        swh_rms:long_name = "RMS of the Ku band significant wave height (from 20
Hz measurements)" ;
        swh_rms:units = "m" ;
```

number of valid points used to compute 1st band significant wave height [swh_num_valid]

```
ubyte swh_num_valid(time) ;
        swh_num_valid:least_significant_digit = 0LL ;
        swh_num_valid:coverage_content_type = "auxiliaryInformation" ;
        swh_num_valid:band = "Ku" ;
        swh_num_valid:coordinates = "lat lon" ;
        swh_num_valid:authority = "CF 1.7" ;
        swh_num_valid:long_name = "number of high frequency valid points used to
compute Ku band significant wave height" ;
        swh_num_valid:units = "1" ;
```

sigma0 [sigma0]

```
double sigma0(time) ;
        sigma0:_FillValue = 1.e+20 ;
        sigma0:comment = "All instrumental corrections included. As available
from retracker and unedited." ;
        sigma0:least_significant_digit = 3LL ;
        sigma0:ancillary_variables = "sigma0_quality" ;
        sigma0:coverage_content_type = "physicalMeasurement" ;
        sigma0:band = "Ku" ;
        sigma0:coordinates = "lat lon" ;
        sigma0:standard_name =
"surface_backwards_scattering_coefficient_of_radar_wave" ;
        sigma0:authority = "CF 1.7" ;
        sigma0:long_name = "Ku band backscatter coefficient" ;
        sigma0:units = "dB" ;
```

sigma0 RMS [sigma0_rms]

```
double sigma0_rms(time) ;
        sigma0_rms:_FillValue = 1.e+20 ;
        sigma0_rms:least_significant_digit = 3LL ;
        sigma0_rms:coverage_content_type = "auxiliaryInformation" ;
        sigma0_rms:band = "Ku" ;
        sigma0_rms:coordinates = "lat lon" ;
        sigma0_rms:authority = "CF 1.7" ;
        sigma0_rms:long_name = "RMS of the high frequency Ku band backscatter
```

```
coefficient" ;
        sigma0_rms:units = "dB" ;
```

number of valid points used to compute backscatter coefficient [sigma0_num_valid]

```
ubyte sigma0_num_valid(time) ;
        sigma0_num_valid:least_significant_digit = 0LL ;
        sigma0_num_valid:coverage_content_type = "auxiliaryInformation" ;
        sigma0_num_valid:band = "Ku" ;
        sigma0_num_valid:coordinates = "lat lon" ;
        sigma0_num_valid:authority = "CF 1.7" ;
        sigma0_num_valid:long_name = "number of high frequency valid points used
to compute Ku band backscatter coefficient" ;
        sigma0_num_valid:units = "1" ;
```

3.1.2.4 SWH quality level and rejection flags

Quality control of individual altimeter measurements is performed with checks over the calculated 1 Hz values and ancillary variables. As a result, the SWH comes with a quality level provided in the *quality_level* variable. Its meaning is defined as follows:

value	meaning	description
0	undefined	the measurement value is not defined or relevant (missing value, etc…), no quality check was applied.
1	bad	the measurement was qualified as not usable after quality check.
2	acceptable	the measurement may still be usable for some application or the quality check could not fully assess if it is a bad or good value (suspect).
3	good	the measurement is usable.

The tests performed to set SWH measurements to a specific quality level are provided below:

meaning	reasons
undefined	• the 1 Hz SWH or SWH RMS could not be calculated (the number of remaining 20Hz SWH after all check and outlier tests is equal to zero)
bad	• the number of valid 20Hz SWH measurements averaged to compute the 1 Hz value is lower than the selected threshold (12 for Saral, 6 for all other

	altimeters) ● the 1 Hz SWH value is not in]0., 30.] meter range ● the 1 Hz SWH RMS is equal to 0. or NaN ● sea ice concentration is greater than 10% ● the 1 Hz location is on land (using OpenstreeMap coastline, 50m precision). Since the 1 Hz location is an average of 20 Hz measurements, this may include mixed land/sea areas. ● the 1 Hz SWH value did not pass the SWH RMS test (described below) ● the 1 Hz SWH value did not pass the outlier test
acceptable	● sea ice concentration is non zero but lower than 10%
good	● all remaining measurements

The quality flag is provided in the *swh_quality* variable:

significant wave height quality [swh_quality]

```
ubyte swh_quality(time) ;
        swh_quality:least_significant_digit = 0LL ;
        swh_quality:flag_values = 0LL, 1LL, 2LL, 3LL ;
        swh_quality:flag_meanings = "undefined bad acceptable good" ;
        swh_quality:coverage_content_type = "qualityInformation" ;
        swh_quality:band = "Ku" ;
        swh_quality:coordinates = "lat lon" ;
        swh_quality:authority = "CF 1.7" ;
        swh_quality:long_name = "quality of Ku band significant wave height
measurement" ;
```

When SWH measurements were rejected as **bad**, the reason (quality test) for which they were rejected is reported in the related *swh_rejection_flags* variable. The following table provides the meaning of each flag possibly raised:

flag	**criteria**
not_water	the surface type is not water. It may be land, continental ice,.... We try to keep lake and inner seas measurements (when the discrimination is possible from the GDR information)
sea_ice	the measurement has possible ice contamination. The sea ice fraction is taken from an external source (such as the CCI Sea Ice microwave based daily maps). Sea ice contamination is defined as

	areas where the sea ice fraction is greater than a minimal threshold (corresponding to 10% of ice in the current configuration).
swh_validity	the SWH measurements were considered as invalid (out of the]0, 30] meter range).
sigma0_validity	the sigma0 measurements were considered as invalid for water surface type. A placeholder, as this not tested in version 2.
waveform_validity	the measurements were considered as invalid as there are indications of unsuitable waveforms for a proper SWH calculation. In particular there was no remaining 20 Hz values after all checks and outlier tests.
ssh_validity	not used.
swh_rms_outlier	the measurements were considered as invalid when the RMS of the SWH measurements used to estimate each 1 Hz SWH measurement was beyond the acceptable threshold for a given range of SWH.
swh_outlier	the measurements were considered as invalid when performing the SWH outlier test, based on the neighbouring measurements within a 100 km window.

Rejection flags [rejection_flags]

```
ushort swh_rejection_flags(time) ;
        swh_rejection_flags:least_significant_digit = 0LL ;
        swh_rejection_flags:flag_meanings = "not_water sea_ice swh_validity
sigma0_validity waveform_validity ssh_validity swh_rms_outlier swh_outlier" ;
        swh_rejection_flags:flag_masks = 1LL, 2LL, 4LL, 8LL, 16LL, 32LL, 64LL,
128LL ;
        swh_rejection_flags:coverage_content_type = "qualityInformation" ;
        swh_rejection_flags:band = "Ku" ;
        swh_rejection_flags:coordinates = "lon lat" ;
        swh_rejection_flags:authority = "CF 1.7" ;
        swh_rejection_flags:long_name = "consolidated instrument and ice flags"
;
        swh_rejection_flags:rms_threshold_lut = "cci_swh_rms_LUT_jason-2.dat" ;
```

SWH rms test

Following Sepulveda et al. (2015), the root mean square deviation of the 20Hz SWH records from the median (1 Hz) swh (swh_rms) is used to reject erroneous swh in the version 2 of the Sea State CCI L2 products.

For each mission included in the Sea State CCI dataset, 90 days of 1Hz SWH measurements were considered, and invalid SWH values associated with the “not_water”, “sea_ice”, “swh_validity” or “waveform_validity” flags were removed from the sample. In order to derive an upper swh rms threshold that is sea state dependent, the mean and standard deviation of the swh rms were computed for SWH bins of 0.5 m width, ranging from 0 to 15 m, with a 0.05 m increment. SWH rms values were considered as outliers when they exceeded a swh rms threshold defined as:

swm_rms_thresh = median(swh_rms)+3*std(swh_rms).

In order to reduce discontinuity in the swh_rms_thresh function for large SWH values, where the number of records is too loo to derive robust statistics, a second-order polynomial function was fitted to the swh_rms_thresh data for SWH values comprised between 3 and 7m. A look-up table of swh rms thresholds was then derived as the swh_rms_thresh values for swh lower than 3m and the values of the polynomial function for swh higher than 3m. For swh above 12 m, a constant value equal to the swh rms threshold at 12m was taken.

Sepulveda, H., Queffeulou, P., Ardhuin, F., 2015. Assessment of Saral/AltiKa Wave Height Measurements Relative to Buoy, Jason-2, and Cryosat-2 Data. Marine Geodesy 38, 449–465. https://doi.org/10.1080/01490419.2014.1000470

3.1.2.5 Adjusted SWH

Following the methodology developed for the version 1 dataset (see Dodet et al., 2020), we have selected the Jason-2 mission as reference to intercalibrate the remaining missions. A first step is to calibrate this reference mission against in situ data, considered as the ground truth. In a second step, the remaining missions (namely, Envisat, Jason-1, Saral, CryoSat-2 and Jason-3) are compared against the calibrated Jason-2 data at crossover locations in order to perform inter-calibration. Finally, an independent verification of this two-step calibration methodology is performed by comparing the statistical metrics between altimeter and in situ data before and after the calibration corrections are applied.

The adjusted significant wave height is provided in *swh_adjusted* variable:

adjusted significant wave height [swh_adjusted]
`double swh_adjusted(time) ;` `swh_adjusted:_FillValue = 1.e+20 ;`

```
        swh_adjusted:comment = "All instrumental corrections included. Adjusted
and unedited" ;
        swh_adjusted:least_significant_digit = 3LL ;
        swh_adjusted:ancillary_variables = "swh_quality swh_rejection_flags" ;
        swh_adjusted:coverage_content_type = "physicalMeasurement" ;
        swh_adjusted:band = "Ku" ;
        swh_adjusted:coordinates = "lat lon" ;
        swh_adjusted:standard_name = "sea_surface_wave_significant_height" ;
        swh_adjusted:authority = "CF 1.7" ;
        swh_adjusted:long_name = "Ku band adjusted significant wave height" ;
        swh_adjusted:units = "m" ;
        swh_adjusted:adjustment_lut = "tab_calibration_lut_jason-1.dat" ;
        swh_adjusted:adjustment_reference = "CCI Sea State v2 Product
Specification Document, 2021" ;
```

Selection of in situ platforms

For the absolute calibration of the altimeter reference mission, we have selected a set of 72 quality controlled in situ platforms from the reprocessed product (WAVE-REP) of the CMEMS INSTAC database (http://www.marineinsitu.eu/). This selection corresponds to offshore platforms (>100km from the coast) with more than 1000 matchups with altimeter data. This selection contains stations mostly from the US National Data Buoy Center (60%), the Canadian Marine Environmental Data Service (15%) and the UK Met Office (8%). The remaining stations (17%) come from other US and European organisms.

Data pairing method

Match-ups : A match-up between altimeter and in situ measurements is identified when the closest approach of the altimeter ground track to the in situ platform location is less than 100 km, within a 30-minute time window. The altimeter match-up data is estimated as the 50-km average of the along-track swh in order to filter out high-frequency variability.

Cross-overs : A cross-over between two altimeter missions is identified when the two altimeter ground tracks cross each other within a one-hour time window. The altimeter cross-over data are estimated as the 50-km average of the along-track Hs in order to filter out high-frequency variability.

Tandem phases: During tandem phases (Jason-1/Jason-2 and Jason-2/Jason-3), the two altimeter missions are sharing the same orbit with a slight time delay (less than 1 min) and the so-called “cross-over” data simply correspond to the swh data recorded at the same location, after applying a 50-km along-track moving average.

Calibration method

In order to account for the non-linear error characteristics of radar altimeter measurements, in particular for low sea state conditions, the calibration methodology implemented in v2 uses data binning technique, so that the calibration correction can accommodate for swh- and mission-dependent error structure. First, the median of the residuals between uncorrected and reference data (the reference data being in situ match-ups for Jason-2 calibration and the calibrated Jason-2 data at cross-over for the remaining missions) is computed for 0.20m bin width, with a 0.05m increment over the full Hs range. Then, a robust linear regression is fitted through the median of the residuals over the 2.5-6m Hs range to model the error function above 2.5 m. These threshold and range have been estimated in order to ensure a smooth transition between the non-linear corrections for low-to-medium sea state conditions and the linear corrections for medium-to-high sea state conditions. Finally a 5-point moving average is applied to the corrections in order to reduce small scale oscillations.

Verification

In order to validate the calibration method, statistical values (bias, root-mean-squared error, scatter index, correlation coefficient) are computed for each mission (Tables 1 and 2). These results are computed before and after the calibration is performed, using either the reference measurements used for the calibration, or independent measurements.

Statistical metrics computed from comparisons at crossovers

The table below summarizes the statistical metrics for the five altimeter missions inter-calibrated against Jason-2 from the comparisons at cross-over data pairs. These statistics were computed after outliers removal, using only data pairs with a weight > 0 in an iteratively reweighted least-squares robust regression.

Table 3.1. Statistical results based on comparisons at missions crossovers

Mission	#values	Bias (m)		NRMSE (%)		SI (%)		R^2	
		No corr.	Corr.	No corr.	Corr.	No corr.	Corr.	No corr.	Corr.
Envisat	10106	0.06	0	4.52	3.75	3.71	3.37	0.99	0.99
CryoSat-2	15005	-0.14	0	5.59	3.26	3.15	2.97	0.99	0.99
Saral	14800	0.24	0	9.39	3.96	4.72	3.54	0.99	0.99
Jason-1	200097	0.06	0	7.17	6.38	5.93	5.56	0.99	0.99
Jason-3	222995	0.03	0	5.74	5.37	5.15	4.91	0.98	0.99
ALL	92601	0.11	0	6.48	4.54	4.53	04.07	0.99	0.99

Statistical metrics computed from comparison at buoy matchups

Once the altimeter swh records have been corrected, an independent verification of the methodology is carried out by computing error statistics between in situ platform and altimeter records before and after correction. These statistics were computed after outliers removal, using only data pairs with a weight > 0 in an iteratively reweighted least-squares robust regression. Table 2 shows the results for each mission, before and after correction.

Table 3.2.Statistical results based on comparisons at buoy matchups

Mission	**#values**	**Bias (m)**		**NRMSE (%)**		**SI (%)**		**R^2**	
		No corr.	Corr.	No corr.	Corr.	No corr.	Corr.	No corr.	Corr.
Envisat	14131	0.07	-0.01	10.6	9.8	8.8	8.47	0.97	0.97
CryoSat-2	795	-0.09	0.02	09.07	08.04	7.29	7.18	0.97	0.97
Saral	10909	0.24	0.02	14.29	9.22	8.35	7.94	0.97	0.97
Jason-1	11433	0.05	-0.01	11.26	10.28	9.55	8.87	0.96	0.97
Jason-3	7043	0.07	0.02	10.88	9.45	8.98	8.11	0.97	0.98
Jason-2	17146	0.07	0	11.12	9.7	9.24	8.34	0.97	0.97
ALL	10243	0.1	0.01	11.2	9.41	8.7	8.15	0.97	0.97

Overall, we note a clear reduction of the difference between the altimeter and in situ measurements once the calibration is applied. In particular, the mean bias decreases from 0.10m to 0.01m, and the bias of each individual mission is lower than 0.03m, which will ensure a better long-term consistency of the multi-mission merged altimeter dataset.

3.1.2.6 SWH uncertainties

Swh uncertainty was estimated from comparisons between altimeter and in-situ platform measurements. Standard deviations of the difference (sdd) between swh from the altimeter and from the buoy were computed over 0.5m swh bins in order to estimate the swh dependency of the sdd. Ordinary least square regression was then fitted to the sdd, using only bins containing more than 100 values.

For each mission, the swh uncertainty is computed as the prediction interval at the 95% confidence level, which is: swh_uncertainty = 1.96 * A1 * swh + A0, where A1 and A0 are the slope and offset of the regression line fitted through sdd as a function of swh.

Table 3.3. Slope and offsets of the regression line fitted through the standard deviation of the difference between swh from the altimeter and from the buoy compuyted over 0.5m swh bins

Mission	#values	A1	A0
Envisat	14131	0.056	0.079
CryoSat-2	795	0.058	0.040
Saral	10909	0.049	0.078
Jason-1	11433	0.054	0.095
Jason-3	7043	0.048	0.087
Jason-2	17146	0.048	0.101

The uncertainty is provided in *swh_uncertainty* variable:

significant wave height uncertainty [swh_uncertainty]

```
double swh_uncertainty(time) ;
        swh_uncertainty:_FillValue = 1.e+20 ;
        swh_uncertainty:comment = "Standard error calculated from buoy
colocations" ;
        swh_uncertainty:least_significant_digit = 3LL ;
        swh_uncertainty:coverage_content_type = "qualityInformation" ;
        swh_uncertainty:coordinates = "lat lon" ;
        swh_uncertainty:authority = "CF 1.7" ;
        swh_uncertainty:long_name = "best estimate of significant wave height
standard error" ;
        swh_uncertainty:units = "m" ;
        swh_uncertainty:formula = "1.96 * 0.054 * SWH + 0.095" ;
        swh_uncertainty:reference = "CCI Sea State v2 Product Specification
Document, 2021" ;
```

3.1.2.7 Denoised SWH

A non-parametric denoising method based on Empirical Mode Decomposition (EMD, Huang et al., 1998) and inspired by wavelet thresholding is applied to the parameter swh_adjusted (see Kopsinis and McLaughlin, 2009, Quilfen et al., 2018 and Quilfen and Chapron, 2019ab). A detailed description of the method can be found in Quilfen and Chapron, 2019b.

References:

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 454, 903–995. https://doi.org/10.1098/rspa.1998.0193

Kopsinis, Y., McLaughlin, S., 2009. Development of EMD-Based Denoising Methods Inspired by Wavelet Thresholding. IEEE Transactions on Signal Processing 57, 1351–1362. https://doi.org/10.1109/TSP.2009.2013885

Quilfen, Y., Yurovskaya, M., Chapron, B., Ardhuin, F., 2018. Storm waves focusing and steepening in the Agulhas current: Satellite observations and modeling. Remote Sensing Of Environment 216, 561–571. https://doi.org/10.1016/j.rse.2018.07.020

Quilfen, Y., and Chapron, B., 2019. Ocean Surface Wave-Current Signatures From Satellite Altimeter Measurements. Geophysical Research Letters 46, 253–261. https://doi.org/10.1029/2018GL081029

Quilfen Y., Chapron B. (2020). On denoising satellite altimeter measurements for high-resolution geophysical signal analysis. *Advances in Space Research, 68.* https://doi.org/10.1016/j.asr.2020.01.005

Denoised significant wave height [swh_denoised]

```
double swh_denoised(time) ;
        swh_denoised:_FillValue = 1.e+20f ;
        swh_denoised:long_name = "Ku band denoised significant wave height" ;
        swh_denoised:standard_name = "sea_surface_wave_significant_height" ;
        swh_denoised:units = "m" ;
        swh_denoised:coordinates = "time lon lat" ;
        swh_denoised:comment = "All instrumental corrections included. adjusted,
denoised with EMD and unedited" ;
        swh_denoised:band = "Ku" ;
               swh_denoised:coverage_content_type = "physicalMeasurement" ;
```

uncertainty attached to adjusted significant wave height [swh_denoised_uncertainty]

```
float swh_denoised_uncertainty(time) ;
        swh_denoised_uncertainty:_FillValue = 1.e+20f ;
        swh_denoised_uncertainty:least_significant_digit = 3LL ;
        swh_denoised_uncertainty:comment = "EMD denoising by Quilfen et al.
applied to swh_adjusted variable" ;
        swh_denoised_uncertainty:ancillary_variables = "swh_quality
swh_rejection_flags" ;
        swh_denoised_uncertainty:coverage_content_type = "physicalMeasurement" ;
        swh_denoised_uncertainty:band = "Ku" ;
        swh_denoised_uncertainty:coordinates = "lon lat" ;
        swh_denoised_uncertainty:authority = "CF 1.7" ;
        swh_denoised_uncertainty:long_name = "uncertainty attached to
swh_adjusted" ;
        swh_denoised_uncertainty:units = "m" ;
        swh_denoised_uncertainty:adjustment_lut =
"tab_calibration_lut_jason-1.dat" ;
        swh_denoised_uncertainty:adjustment_reference = "CCI Sea State v2
Product Specification Document, 2021" ;
```

high-frequency noise attached to adjusted significant wave height [swh_noise]

float swh_noise(time) ;
        swh_noise:_FillValue = 1.e+20f ;
        swh_noise:least_significant_digit = 3LL ;
        swh_noise:comment = "EMD denoising by Quilfen et al. applied to swh_adjusted variable" ;
        swh_noise:ancillary_variables = "swh_quality swh_rejection_flags" ;
        swh_noise:coverage_content_type = "physicalMeasurement" ;
        swh_noise:band = "Ku" ;
        swh_noise:coordinates = "lon lat" ;
        swh_noise:authority = "CF 1.7" ;
        swh_noise:long_name = "high-frequency noise attached to swh_adjusted" ;
        swh_noise:units = "m" ;
        swh_noise:adjustment_lut = "tab_calibration_lut_jason-1.dat" ;
        swh_noise:adjustment_reference = "CCI Sea State v2 Product Specification Document, 2021" ;

3.1.2.8 Altimeter wind speed

The provided altimeter wind speed is taken from the agency's (S)GDR product, when available, without any adjustment or correction.

Altimeter wind speed [wind_speed_alt]

double wind_speed_alt(time) ;
        wind_speed_alt:least_significant_digit = 3LL ;
        wind_speed_alt:_FillValue = 1.e+20 ;
        wind_speed_alt:long_name = "wind speed from Ku band altimeter, as in GDR" ;
        wind_speed_alt:standard_name = "wind_speed" ;
        wind_speed_alt:authority = "CF 1.7" ;
        wind_speed_alt:units = "m s-1" ;
        wind_speed_alt:ancillary_variables = "wind_speed_alt_quality" ;
        wind_speed_alt:coverage_content_type = "physicalMeasurement" ;
        wind_speed_alt:band = "Ku" ;
        wind_speed_alt:coordinates = "time lon lat" ;

3.1.2.9 Model data

The selected source for model is the ERA5 reanalysis (https://climate.copernicus.eu/climate-reanalysis) : currently it is only filled in from 2000 to

2018, but it is being extended backward in a later version. The parameters currently extracted are:

- U and V wind components at 10m
- air temperature at 2m
- sea level surface pressure

Sea surface temperature and total liquid water content are currently missing and will be added in a later version.

Model wind speed zonal component [wind_speed_model_u]
double wind_speed_model_u(time) ; wind_speed_model_u:_FillValue = 1.e+20f ; wind_speed_model_u:long_name = "U component of the model wind vector" ; wind_speed_model_u:standard_name = "eastward_wind" ; wind_speed_model_u:units = "m s-1" ; wind_speed_model_u:source = "atmospheric model ERA5" ; wind_speed_model_u:institution = "ECMWF" ; wind_speed_model_u:coordinates = "time lon lat" ; wind_speed_model_u:coverage_content_type = "modelResult" ;

Model wind speed meridional component [wind_speed_model_v]
double wind_speed_model_v(time) ; wind_speed_model_v:_FillValue = 1.e+20f ; wind_speed_model_v:long_name = "V component of the model wind vector" ; wind_speed_model_v:standard_name = "northward_wind" ; wind_speed_model_v:units = "m s-1" ; wind_speed_model_v:source = "atmospheric model ERA5" ; wind_speed_model_v:institution = "ECMWF" ; wind_speed_model_v:coordinates = "time lon lat" ; wind_speed_model_v:coverage_content_type = "modelResult" ;

Sea surface temperature [sea_surface_temperature]
double sea_surface_temperature(time) ; sea_surface_temperature:_FillValue = 1.e+20f ; sea_surface_temperature:long_name = "sea surface temperature" ; sea_surface_temperature:standard_name = "sea_surface_temperature" ; sea_surface_temperature:units = "K" ; sea_surface_temperature:source = "atmospheric model ERA5" ; sea_surface_temperature:institution = "ECMWF" ; sea_surface_temperature:coordinates = "time lon lat" ; sea_surface_temperature:coverage_content_type = "modelResult" ;

Surface air temperature [surface_air_temperature]
double surface_air_temperature(time) ; surface_air_temperature:_FillValue = 1.e+20f ; surface_air_temperature:long_name = "surface air temperature" ; surface_air_temperature:standard_name = "air_temperature" ; surface_air_temperature:units = "K" ; surface_air_temperature:source = "atmospheric model ERA5" ; surface_air_temperature:institution = "ECMWF" ; surface_air_temperature:coordinates = "time lon lat" ; surface_air_temperature:coverage_content_type = "modelResult" ;

Surface air pressure [surface_air_pressure]
double surface_air_pressure(time) ; surface_air_pressure:_FillValue = 1.e+20f ; surface_air_pressure:long_name = "surface air pressure" ; surface_air_pressure:standard_name = "air_pressure_at_mean_sea_level" ; surface_air_pressure:units = "Pa" ; surface_air_pressure:source = "atmospheric model ERA5" ; surface_air_pressure:institution = "ECMWF" ; surface_air_pressure:coordinates = "time lon lat" ; wind_speed_model_u:coverage_content_type = "modelResult" ;

Total column liquid water [total_column_liquid_water_content]
double total_liquid_water_content(time) ; total_liquid_water_content:_FillValue = 1.e+20f ; total_liquid_water_content:long_name = "total column liquid water" ; total_liquid_water_content:standard_name = "atmosphere_cloud_liquid_water_content" ; total_liquid_water_content:units = "kg m-2" ; total_liquid_water_content:source = "atmospheric model ERA5" ; total_liquid_water_content:institution = "ECMWF" ; total_liquid_water_content:coordinates = "time lon lat" ; total_liquid_water_content:coverage_content_type = "modelResult" ;

3.1.2.10 Sea Surface Height

The sea surface height is taken from the CMEMS Reprocessed Sea Level dataset SEALEVEL_GLO_PHY_L3_REP_OBSERVATIONS_008_062 (instead of CCI Sea Level product in dataset v1, which had a too limited coverage and was missing several recent missions), available for all missions. The following variables are copied from CMEMS product:

Mean dynamic topography [mdt]

```
float mdt(time) ;
        mdt:_FillValue = 1.e+20f ;
        mdt:comment = "The mean dynamic topography is the sea surface height
above geoid; it is used to compute the absolute dynamic tyopography adt=sla+mdt"
;
        mdt:long_name = "Mean dynamic topography" ;
        mdt:standard_name = "sea_surface_height_above_geoid" ;
        mdt:units = "m" ;
        mdt:source = "SEALEVEL_GLO_PHY_L3_REP_OBSERVATIONS_008_062" ;
        mdt:institution = "CMEMS" ;
        mdt:authority = "CF-1.7" ;
        mdt:coverage_content_type = "auxiliaryInformation" ;
        mdt:source_files =
"dt_global_j1n_phy_l3_20100719_20190101.ncdt_global_j1n_phy_l3_20100720_20190101
.nc" ;
        mdt:coordinates = "lat lon" ;
```

Sea level anomaly [sla_unfiltered]

```
float sla_unfiltered(time) ;
        sla_unfiltered:_FillValue = 1.e+20f ;
        sla_unfiltered:comment = "The sea level anomaly is the sea surface
height above mean sea surface height; the uncorrected sla can be computed as
follows: [uncorrected sla]=[sla from product]+[dac]+[ocean_tide]-[lwe]; see the
product user manual for details" ;
        sla_unfiltered:long_name = "Sea level anomaly not-filtered
not-subsampled with dac, ocean_tide and lwe correction applied" ;
        sla_unfiltered:standard_name = "sea_surface_height_above_sea_level" ;
        sla_unfiltered:units = "m" ;
        sla_unfiltered:source = "SEALEVEL_GLO_PHY_L3_REP_OBSERVATIONS_008_062" ;
        sla_unfiltered:institution = "CMEMS" ;
        sla_unfiltered:authority = "CF-1.7" ;
        sla_unfiltered:coverage_content_type = "auxiliaryInformation" ;
        sla_unfiltered:source_files =
"dt_global_j1n_phy_l3_20100719_20190101.ncdt_global_j1n_phy_l3_20100720_20190101
.nc" ;
        sla_unfiltered:coordinates = "lat lon" ;
```

3.1.2.11 Sea ice concentration

Sea ice concentration is provided as an ancillary variable but also used in the SWH editing procedure. We use an external sea ice concentration product to discard possibly ice contaminated measurements. Because no products provides a complete temporal coverage (missed acquisitions, non continuities between different microwave radiometer missions, infrequent updates of some datasets), we had to use three different sources, which are in order of priority:

1. CCI Sea Ice Concentration from AMSR, v2.1 (2002-2017)
2. OSI SAF Sea Ice Concentration CDR reprocessing v2p0 (1979-2015)
3. OSI SAF Sea Ice Concentration ICDR reprocessing v2p0 (2016-2021)

For each source, we use the closest in time concentration map, up to three days apart in case it is missing for a given day, before switching to the next source in line. The sea ice concentration is expressed as a fraction.

Sea-ice concentration [sea_ice_fraction]

```
double sea_ice_fraction(time) ;
        sea_ice_fraction:_FillValue = 1.e+20 ;
        sea_ice_fraction:least_significant_digit = 2LL ;
        sea_ice_fraction:coverage_content_type = "auxiliaryInformation" ;
        sea_ice_fraction:coordinates = "lat lon" ;
        sea_ice_fraction:standard_name = "sea_ice_fraction" ;
        sea_ice_fraction:authority = "CF 1.7" ;
        sea_ice_fraction:long_name = "sea ice fraction" ;
        sea_ice_fraction:units = "1" ;
```

3.1.2.12 Bathymetry

The same bathymetry source was used for all mission to get the ocean sea floor depth. We selected the 30 second arc General Bathymetric Chart of the Oceans (GEBCO), 2014 [doi:10.1002/2015EA000107], available at:

https://www.gebco.net/data_and_products/gridded_bathymetry_data/gebco_30_second_grid

Ocean depth [bathymetry]

```
float bathymetry(time) ;
        bathymetry:_FillValue = 1.e+20f ;
        bathymetry:coverage_content_type = "auxiliaryInformation" ;
        bathymetry:source = "The GEBCO_2014 Grid, version 20150318,
www.gebco.net,  doi:10.1002/2015EA000107" ;
        bathymetry:institution = "IOC/IHO" ;
        bathymetry:standard_name = "s" ;
        bathymetry:authority = "CF-1.7" ;
```

```
        bathymetry:long_name = "ocean depth" ;
        bathymetry:units = "m" ;
        bathymetry:coordinates = "lat lon" ;
```

3.1.2.13 Distance to coast

The distance to the nearest coastline for each ocean measurement was extracted from the Distance to Nearest Coastline grid at 0.01 degree resolution, provided by the NASA Goddard Space Flight Center (GSFC) Ocean Color Group and available at:

http://www.pacioos.hawaii.edu/metadata/dist2coast_1deg.html

Distance to nearest coast [distance_to_coast]
double distance_to_coast(time) ; distance_to_coast:_FillValue = 1.e+20 ; distance_to_coast:long_name = "distance to nearest coast" ; distance_to_coast:authority = "CF-1.7" ; distance_to_coast:units = "m" ; distance_to_coast:source = "Distance to Nearest Coastline: 0.01-Degree Grid, by NASA Goddard Space Flight Center (GSFC) Ocean Color Group" ; distance_to_coast:coverage_content_type = "auxiliaryInformation" ; distance_to_coast:institution = "NASA/GFSC" ; distance_to_coast:coordinates = "time lon lat" ;

3.1.3 Planned improvements

Future releases of CCI Sea State dataset will be published in the coming months and years, including the following improvements:

- extending the series to now and filling in existing data gaps (such as the graveyard orbit phases for Jason-1 and Jason-2)
- adding SRAL altimeters onboard Sentinel-3A & 3B
- re-introducing older missions such as ERS-1, ERS-2, GFO and Topex
- improving the editing of SWH data, in particular over sea ice areas, slicks and strong wind areas
- adding the mean square slope
- adding cross-mission adjustment of sigma0
- adding an adjusted altimeter wind speed

3.2 L3

The main objective of the L3 is to provide a lighter, multi-mission, version of the L2P, easier to use for non-expert users.

The L3 dataset consists in the concatenation into daily files of the 1 Hz measurements taken from all available L2P files of all satellite missions available in the version 2 CCI Sea State Dataset. Only the best quality measurements are kept (*swh_quality* = 3). Only a subset of the variables available in L2P is copied into the L3, as listed in the table below.

This section describes the specific variables contained in the L3 of version 2 CCI Sea State Dataset.

coordinate variables	
time	time (in seconds since 1981-01-01)
lat	latitude
lon	longitude
instrumental variables for 1st band altimeter (usually Ku)	
sigma0	Ku band backscatter coefficient, as calculated from the retracking
environmental variables for 1st band altimeter (usually Ku)	
swh	significant wave height, within 1Hz cells, averaged over the 20 Hz measurements calculated from the retracking
swh_adjusted	adjusted significant wave height. The adjustment is based on cross-mission intercalibration as described in section 4.1.2.5.
swh_denoised	EMD-filtered significant wave height, an adjusted and denoised significant wave height estimated by CCI Sea State project and based on Quilfen et al. [Quilfen, Y., and B. Chapron (2019). On denoising satellite altimeter measurements for geophysical signals analysis. Advances in Space Research. Submitted]
swh_uncertainty	best estimate of significant wave height standard error, as estimated by CCI Sea State project for each mission and described in section 4.1.2.6.

topography variables	
bathymetry	ocean depth, taken from GEBCO version 20150318
distance_to_coast	distance to nearest coast, from NASA/GFSC distance to coastline
origin of measurements	
satellite	identifier of the satellite from which the measurement comes from
relative_pass_number	relative pass number of the track which the measurement comes from
cycle_number	cycle number of the track which the measurement comes from

Significant wave height [swh]

```
double swh(time) ;
        swh:_FillValue = 1e20 ;
        swh:long_name = "Ku band significant wave height" ;
        swh:standard_name = "sea_surface_wave_significant_height" ;
        swh:units = "m" ;
        swh:ancillary_variables = "swh_quality rejection_flag" ;
        swh:coordinates = "time lon lat" ;
        swh:comment = "All instrumental corrections included. Unadjusted and
unedited" ;
        swh:band = "Ku" ;
        swh:coverage_content_type = "physicalMeasurement" ;
```

adjusted significant wave height [swh_adjusted]

```
double swh_adjusted(time) ;
        swh_adjusted:_FillValue = 1e20 ;
        swh_adjusted:long_name = "Ku band adjusted significant wave height" ;
        swh_adjusted:standard_name = "sea_surface_wave_significant_height" ;
        swh_adjusted:units = "m" ;
        swh_adjusted:calibration_formula = "1.0149*swh + 0.0277" ;
        swh_adjusted:calibration_reference = "Ash E R & Carter D J T, September
2010, Satellite wave data quality report, GlobWave Deliverable D.16" ;
```

```
        swh_adjusted:ancillary_variables = "swh_quality swh_uncertainty
rejection_flag" ;
        swh_adjusted:coordinates = "time lon lat" ;
        swh_adjusted:comment = "All instrumental corrections included. adjusted
and unedited" ;
        swh_adjusted:band = "Ku" ;
        swh_adjusted:coverage_content_type = "physicalMeasurement" ;
```

Denoised significant wave height [swh_denoised]

```
double swh_denoised(time) ;
        swh_denoised:_FillValue = 1e20 ;
        swh_denoised:long_name = "Ku band denoised significant wave height" ;
        swh_denoised:standard_name = "sea_surface_wave_significant_height" ;
        swh_denoised:units = "m" ;
        swh_denoised:coordinates = "time lon lat" ;
        swh_denoised:comment = "All instrumental corrections included. adjusted,
denoised with EMD and unedited" ;
        swh_denoised:band = "Ku" ;
        swh_denoised:coverage_content_type = "physicalMeasurement" ;
```

significant wave height standard deviation [swh_std]

```
double swh_uncertainty(time) ;
        swh_uncertainty:_FillValue = 1e20  ;
        swh_uncertainty:long_name = "standard deviation of the Ku band
significant wave height" ;
        swh_uncertainty:units = "m" ;
        swh_uncertainty:coordinates = "time lon lat" ;
        swh_uncertainty:coverage_content_type = "auxiliaryInformation" ;
```

sigma0 [sigma0]

```
double sigma0(time) ;
        sigma0:_FillValue = 1e20 ;
        sigma0:long_name = "Ku band backscatter coefficient" ;
        sigma0:band = "Ku" ;
        sigma0:units = "dB" ;
        sigma0:quality_flag = "sigma0_quality" ;
        sigma0:coordinates = "time lon lat" ;
        sigma0:comment = "All instrumental corrections included. Unadjusted" ;
        sigma0:coverage_content_type = "physicalMeasurement" ;
```

Ocean depth [bathymetry]

```
float bathymetry(time) ;
        bathymetry:_FillValue = 1.e+20f ;
        bathymetry:coverage_content_type = "auxiliaryInformation" ;
        bathymetry:source = "The GEBCO_2014 Grid, version 20150318,
www.gebco.net,  doi:10.1002/2015EA000107" ;
        bathymetry:institution = "IOC/IHO" ;
        bathymetry:standard_name = "s" ;
        bathymetry:authority = "CF-1.7" ;
        bathymetry:long_name = "ocean depth" ;
        bathymetry:units = "m" ;
        bathymetry:coordinates = "lat lon" ;
```

Distance to nearest coast [distance_to_coast]

```
double distance_to_coast(time) ;
        distance_to_coast:_FillValue = 1.e+20 ;
        distance_to_coast:long_name = "distance to nearest coast" ;
        distance_to_coast:authority = "CF-1.7" ;
        distance_to_coast:units = "m" ;
        distance_to_coast:source = "Distance to Nearest Coastline: 0.01-Degree
Grid, by NASA Goddard Space Flight Center (GSFC) Ocean Color Group" ;
        distance_to_coast:coverage_content_type = "auxiliaryInformation" ;
        distance_to_coast:institution = "NASA/GFSC" ;
        distance_to_coast:coordinates = "time lon lat" ;
```

satellite associated with the measurement [satellite]

```
ubyte satellite(time) ;
        satellite:long_name = "satellite associated with measurement" ;
        satellite:flag_values = 0UB, 1UB, 2UB, 3UB, 4UB, 5UB, 6UB, 7UB, 7UB,
8UB, 9UB, 10UB ;
        satellite:flag_meanings = "cryosat-2 jason-1 jason-2 jason-3 saral
sentinel-3_a envisat topex topex-poseidon ers-1 ers-2 gfo" ;
        satellite:coverage_content_type = "auxiliaryInformation" ;
        satellite:coordinates = "lon lat" ;
```

relative pass number of satellite track associated with measurement [relative_pass_number]

```
ushort relative_pass_number(time) ;
        relative_pass_number:_FillValue = 0US ;
```

```
        relative_pass_number:coverage_content_type = "auxiliaryInformation" ;
        relative_pass_number:long_name = "relative pass number of satellite
associated with measurement" ;
        relative_pass_number:coordinates = "lon lat" ;
```

cycle number of satellite associated with measurement [cycle_number]

```
ushort cycle_number(time) ;
        cycle_number:coverage_content_type = "auxiliaryInformation" ;
        cycle_number:long_name = "cycle number of satellite associated with
measurement" ;
        cycle_number:coordinates = "lon lat"
```

3.3 L4

The main objective of the L4 dataset is to provide monthly composites of simple significant wave height statistics computed from all available missions in version 2 CCI Sea State Dataset, for climatological studies or visualization.

The L4 dataset characteristics are as follow:

- each altimeter track of each mission is gridded independently over a 1°x1° global grid. The measurements of a given altimeter track falling into a given 1°x1° grid cell are averaged using a median operator. Each computed grid cell value is later referred to as a **median significant wave height**. The different L4 statistics are then computed for each grid cell over the individual track's median significant wave height computed for this particular cell.
- the SWH measurements gridded for each mission are the **denoised adjusted significant wave heights** (swh_denoised from L2P datasets).

The following table describes the specific statistics variables contained in the L4 of version 2 CCI Sea State Dataset.

	coordinate variables
time	time (in seconds since 1985-01-01)
lat	latitude
lon	longitude
	environmental variables for 1st band altimeter (usually Ku)
swh_mean	mean of median significant wave height values
swh_rms	rms of median significant wave height values
swh_num	number of median significant wave height values
swh_sum	total of median significant wave height values
swh_squared_sum	total of median significant wave height squared values
swh_log_sum	total of median significant wave height log values

swh_log_squared_sum	total of median significant wave height log squared values
swh_num_gt0050	number of median significant wave height values greater than 0.5m
swh_num_gt0100	number of median significant wave height values greater than 1.0m
swh_num_gt0150	number of median significant wave height values greater than 1.5m
swh_num_gt0200	number of median significant wave height values greater than 2.0m
swh_num_gt0250	number of median significant wave height values greater than 2.5m
swh_num_gt0300	number of median significant wave height values greater than 3.0m
swh_num_gt0350	number of median significant wave height values greater than 3.5m
swh_num_gt0400	number of median significant wave height values greater than 4.0m
swh_num_gt0500	number of median significant wave height values greater than 5.0m
swh_num_gt0600	number of median significant wave height values greater than 6.0m
swh_num_gt0800	number of median significant wave height values greater than 8.0m
swh_num_gt1000	number of median significant wave height values greater than 10.0m
swh_max	maximum median significant wave height value

3.5 Global Attributes

Global Attribute	Example content
time_coverage_start	2008-07-04T12:19:20Z
time_coverage_end	2008-07-04T12:30:39Z
Conventions	CF-1.7, ACDD-1.3, ISO 8601
title	ESA CCI Sea State L2P derived from Jason-2 GDR
id	ESACCI-SEASTATE-L2P-SWH-Jason-2
institution	Institut Francais de Recherche et d\'Exploitation de la Mer/Centre de Recherche et d\'Exploitation satellitaire
institution_abbreviation	ifremer/cersat
source	CCI Sea State Jason-2 GDR to L2P Processor
history	2021-07-20 18:50:45.929774 - Creation
references	CCI Sea State Product Specification Document (PSD), v1.1
product_version	1.0
summary	This dataset contains along-track significant wave height measurements from Jason-2 altimeter, cross-calibrated with other altimetry missions and reference in situ measurements.
keywords	Oceans > Ocean Waves > Significant Wave Height", "Oceans > Ocean Waves > Sea State

keywords_vocabulary	NASA Global Change Master Directory (GCMD) Science Keywords
naming_authority	fr.ifremer.cersat
cdm_data_type	trajectory
featureType	trajectory
comment	These data were produced at ESACCI as part of the ESA SST CCI project.
creator_name	Cersat
creator_url	http://cersat.ifremer.fr
creator_email	cersat@ifremer.fr
creator_institution	Ifremer / Cersat
project	Climate Change Initiative - European Space Agency
geospatial_lat_min	-80.
geospatial_lat_max	80.
geospatial_lat_units	degrees
geospatial_lon_min	-180.
geospatial_lon_max	180
geospatial_lon_units	degrees
standard_name_vocabulary	NetCDF Climate and Forecast (CF) Metadata Convention
license	ESA CCI Data Policy: free and open access
platform	Jason-2

platform_type	low earth orbit satellite
platform_vocabulary	CCI
instrument	Poseidon-3
instrument_type	altimeter
instrument_vocabulary	CCI
spatial_resolution	5.8km
cycle_number	0
relative_pass_number	61
equator_crossing_time	2008-07-04 12:02:33.226000
equator_crossing_longitude	
netcdf_version_id	4.6.1 of May 13 2018 11:35:43
acknowledgement	Please acknowledge the use of these data with the following statement: these data were obtained from the ESA CCI Sea State project
format_version	Data Standards v2.1
processing_level	L2P
track_id	7ae70d21-32b1-4c86-b9d3-8fe7f984bf22
publisher_name	ifremer/cersat
publisher_url	http://cersat.ifremer.fr
publisher_email	cersat@ifremer.fr

publisher_institution	Ifremer / Cersat
scientific_support_contact	Guillaume.Dodet@ifremer.fr
technical_support_contact	cersat@ifremer.fr
key_variables	swh_adjusted_denoised
date_created	2021-07-20T18:50:45
date_modified	2021-07-20T18:50:45
band	Ku
source_version	2.0.6
input	GDR product: JA2_GPS_2PdP000_061_20080704_113426_20080704_123039.nc
processing_software	Cersat/Cerbere 1.0
Metadata_Conventions	Climate and Forecast (CF) 1.7, Attribute Convention for Data Discovery (ACDD) 1.3
geospatial_vertical_units	meters above mean sea level
geospatial_vertical_positive	up

4. SAR Datasets

4.1 Overview of SAR Products

Traditional SAR wave mode spectral inversion is well known to produce a 2D spectrum over a limited azimuthal wavenumber range, the so called azimuth cutoff caused by the surface wave orbital velocity variance within a SAR resolution cell. Therefore, most of the time, only swell wave spectra can be estimated. In this project, an extension of this inversion has been developed by several research groups to be able to estimate the significant wave height of the full wave spectra, including the wind sea. The inputs are also the SAR image intensity and the SAR cross-spectra but the output is the integrated significant wave height parameters, comparable to the one derived from altimeters. The different algorithms are all based on machine learning using a training dataset composed by either colocated wave model or altimeter data.

Different flavors of this algorithm have been developed by several groups, ending up, after successive improvements to a very comparable performance. Two algorithms have been proposed for the Sentinel1 wave mode from a team lead by Brandon Quach and another led by Andrey Pleskachevsky and one algorithm has been proposed for Envisat ASAR wave mode by Li, based on CWAVE developed at DLR.

Intended usage of such products is the same as the significant wave height from altimeters and provides an independent assessment of the SWH parameter long term evolution for climate studies or validation of sea state models, especially using Quach algorithm, that is not using any model data in the training process.

Algorithm	Missions	Temporal coverage	Benefits	Limitations
Quach	Sentinel-1A, Sentinel-1B	1-Dec-2014 to 23-Feb-2021	Independent of model information.	Only swh variable provided.
Pleskachevsky	Sentinel-1A, Sentinel-1B	1-Dec-2014 to 23-Feb-2021	Includes periods and partially integrated wave parameters.	Only trained with model data.
Li and Huang	Envisat	10-Dec-2002 to 8 Apr-2012	Best algorithm currently available for Envisat ASAR WM.	

Table 4.1: Comparison of SAR product algorithms.

4.2 Wave height estimation from Sentinel-1 WV using *Quach et al* methodology (IFREMER & University of Hawaii)

4.2.1 Methodology

University of Hawaii and IFREMER estimate the significant wave height from synthetic aperture radar (SAR) using statistical models. The accuracy of existing models is limited by the reliance on wave spectra from global physics models rather than direct observations to fit parameters. The observational data set consists of 780,000 collocations between five altimeters and two Sentinel-1 (S-1) SAR satellites. The altimeter database used as target reference is the Ribal and Young, (2020) calibrated altimetry dataset. It was used as reference to train the convolutional neural network (CNN). 2015-2017 observations were used to train the model and test the model on data from 2018. The altimetry dataset is the reference to train a deep neural network regression model for predicting significant wave height and its uncertainty. The reference altimeter observations and neural network architecture improve wave height root mean square error (RMSE) by 25 cm with respect to previous paper (Stopa and Mouche 2017), achieving RMSEs of 0.25-0.4 m relative to independent observations from altimeters and buoys. This reduces the error of the current state-of-the-art approach by half. In the round-robin competition the model most accurately predicted the wave height relative to other models such as Stopa and Mouche (2017) and an retrained model of Schulz-Stellenfleth et al., (2007) in Pleskachevsky et al., (2019) which used wave model data and linear regression models or shallow neural networks.

The CNN uses input from the image modulation spectra from multiple looks of an ocean scene. Other features were included in the training dataset: geographical parameters such as the normalized radar cross section, normalized image variance, incidence angle, latitude, longitude, and time of day. Additionally, an orthogonal set of parameters following Schulz-Stellenfleth et al., (2007), called CWAVE has been added for the Hs regression. The CWAVE parameters have minimal impact within the model and as the training data increases the dependence on the CWAVE parameters decreases. The CWAVE parameters have been kept in the model because of the improvement in the extremely small and large wave heights. These parameters help to reduce the effects of over fitting. Ribal and Young, (2020) dataset is showing consistent significant wave height estimations between altimeter platforms. Furthermore, the Sentinel-1 platforms: Sentinel-1 A and Sentinel-1 B are also well calibrated to each other and the data can be used interchangeably. For further details of the model setup see Quach et al., (2021).

Spatial wave height RMSEs are larger in the extra-tropics (0.8 m) relative to the low latitudes (0.2 m). Accurate predictions wave heights can be achieved in nearly all SAR images over the ocean even when the SAR images contain distinct atmospheric imprints. In elevated wave heights (Hs>6 m), the model demonstrates an average wave height accuracy within 0.5-1 m of the altimeter observations in both tropical and extratropical cyclones. Instead of relying on the set of engineered CWAVE features that capture most of the discriminative information, the deep learning approach learns directly from the low-level, high-dimensional image spectra. Furthermore, Quach et al 2021 results indicate that there is still room for improvement with additional training data, especially in extreme sea states with Hs > 8 m.

Thus, one can expect the statistical model to improve as more collocation events are collected.

References:

Pleskachevsky, A., S. Jacobsen, B. Tings, and E. Schwarz, “Estimation of sea state from Sentinel-1 synthetic aperture radar imagery for maritime situation awareness,” Int. J. Remote Sens., vol. 40, no. 11, pp. 4104–4142, Jan. 2019.

Quach, B., Y. Glaser, J. E. Stopa, A. Mouche, P. Sadowski, 2020. Deep Learning for Predicting Significant Wave Height from Synthetic Aperture Radar, IEEE Transactions on Geoscience and Remote Sensing, pgs 1-9, doi:10.1109/TGRS.2020.3003839

Ribal A., and I. R. Young, “33 years of globally calibrated wave height and wind speed data based on altimeter observations,” Scientific Data, vol. 6, no. 1, pp. 1–5, May 2019.

J. E. Stopa and A. Mouche, “Significant wave heights from Sentinel-1 SAR: Validation and applications,” J. Geophys. Res., Oceans, vol. 122, no. 3, pp. 1827–1848, Mar. 2017.

J. Schulz-Stellenfleth, T. König, and S. Lehner, “An empirical approach for the retrieval of integral ocean wave parameters from synthetic aperture radar data,” J. Geophys. Res., vol. 112, Mar. 2007, Art. no. C03019, doi: 10.1029/2006JC003970.

4.2.2 Content of Quach SAR product

	coordinate variables
time	time (in seconds since 1981-01-01)
lat	latitude
lon	longitude
	instrumental variables for SAR
angle_of_incidence	incidence angle of the WV acquisition
heading	satellite heading relative to geographic North in clockwise convention

	environmental variables for SAR
swh	total C band significant wave height
swh_uncertainty	standard deviation associated to hs: level of confidence of the NN model
swh_quality	quality of C band significant wave height measurement
swh_rejection_flags	consolidated instrument and ice flags
wind_speed	wind speed coming from ESA OCN WV product (CMOD-based wind inversion without Bayesian scheme)
sigma0	sigma0 (Normalized Radar Cross Section)
normalized_variance	Normalized variance of digital numbers (complex I+Q values)
	auxiliary variables
auxiliary measurements	
sea_ice_fraction	sea ice fraction, as taken from CCI Sea Ice products
topography	
bathymetry	ocean depth, taken from GEBCO version 20150318
distance_to_coast	distance to nearest coast, from NASA/GFSC distance to coastline
model auxiliary data	
wind_speed_model_u	U component of the model wind vector, taken from ERA5 reanalysis
wind_speed_model_v	V component of the model wind vector, taken from ERA5 reanalysis

surface_air_temperature	surface air temperature, taken from ERA5 reanalysis
surface_air_pressure	surface air pressure, taken from ERA5 reanalysis
swh_model	Copernicus ERA5 Reanalysis by ECMWF

4.2.3 Validation of Quach SAR product

Several inter comparisons have been performed between the Hs estimation using Quach 2020 algorithm and moored buoys, altimeters, numerical model forecasts. The first document to cite is [*Quach et al, 2020*](#) . The validation against SWIM nadir beam onboard CFOSAT mission is particularly interesting because the product provided by CNES (French spatial agency) is not present in the training dataset (Ribal and Young altimeters database) and the quality of the SWH at 1Hz have demonstrated good performances (see Hauser et al, 2020 10.1109/TGRS.2020.2994372) with respect to MF-WAM numerical model. It is worth mentioning that this product is using one of the most recent re-tracker algorithms, the so-called 'Adaptive retracking'.

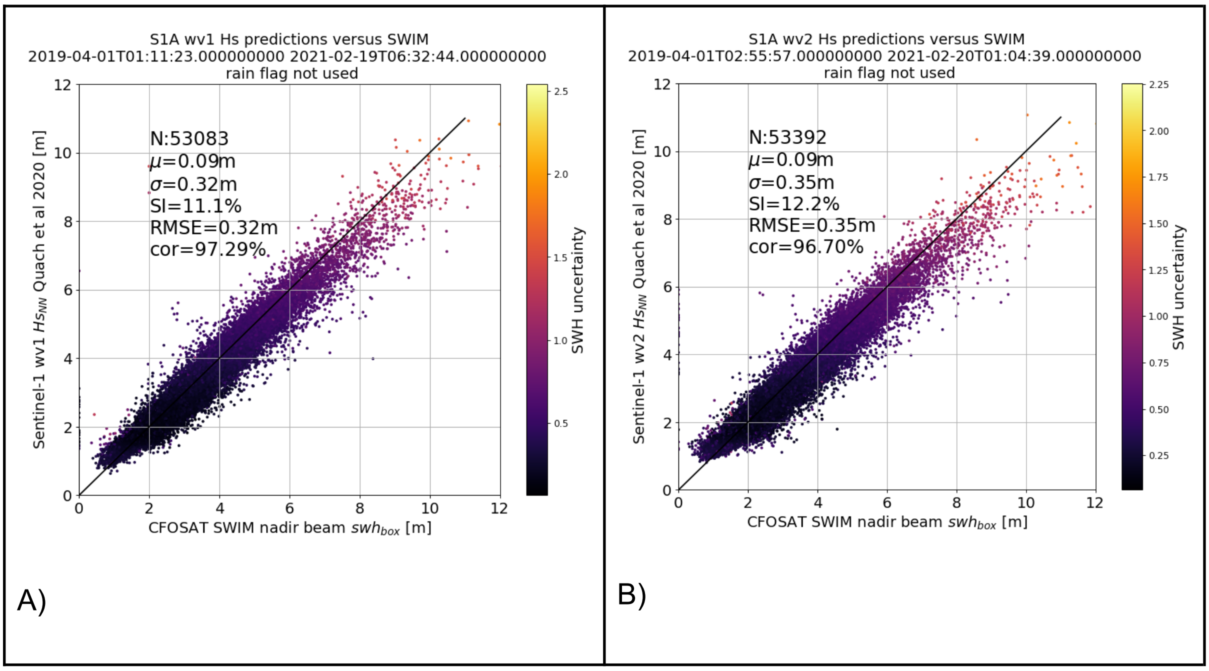

Figure 4.1 A) and B) are examples of SWH performances for WV1 (resp. WV2) with respect to SWIM nadir 1Hz measurements in open ocean with acquisitions in between 55° North and 55° South (to avoid sea-ice or icebergs contaminations).

4.3 Wave height estimation from Sentinel-1 WV using *Pleskachevsky et al.* methodology (DLR)

4.3.1 Methodology

Algorithm Overview

An empirical algorithm was developed for estimation of wide spectrum of integrated sea state parameters from satellite-borne Synthetic Aperture Radar (SAR): total significant wave height, dominant and secondary swell wave height, windsea wave height, first and second moment wave periods, mean wave period and period of windsea (see Table 5.3, Table 5.4). The algorithm was tuned and validated for C-band Sentinel-1 (S1) Wave Mode (wv) products consisting of multiple 20km×20km imagettes. For each imagette, one integrated value has been estimated. The current S1 wv method is an extension of sea state processor developed for sea state parameter estimation from S1 Interferometric Wide Swath Mode (IW) (Pleskachevsky et al., 2019) and EW (Extra Wide) SAR images for Near Real Time applications.

The algorithm consists of two steps for estimation of sea state parameters: the initial step is based on classical CWAVE parameters related by linear regression model function (Schulz-Stellenfleth et al., 2007) and extended with additional features. The second postprocessing step uses machine learning, i.e. the support vector machine technique (SVM), to improve the accuracy of the initial results. In this way, the accuracy of initial SWH estimated using linear regression with ~0.35m accuracy for S1 WV (CMEMS-hindcast validation, Pleskachevsky et al., 2021) can be improved to ~0.25m accuracy by SVM postprocessing .

The combined algorithm is able to process S1 Level-1 SLC and GRD data in near real time (NRT). The complete archive of S1 WV from December 2014 until February 2021 with ~60 overflights/day each including ~120 imagettes was processed. Validation was executed using CMEMS sea state hindcast within latitude of -60°<LAT<60° to avoid ice coverage.

The method was trained and validated with CMEMS 3h (temporally interpolated) model results and additionally validated against NDBC buoys collocated with maximum distance of 50 km. Worldwide were found 61 buoys with significant wave height SWH records for 2015-2021 using this distance.

For SVM postprocessing, the total number of validated imagettes is around 4 Mio for 2018-2021 (2016-2017 data was used for training) and results into RMSE=0.25m for wv1 imagettes (incidence angle of 23°) and RMSE=0.28m for wv2 imagettes (incidence angle of 36°). The comparisons with NDBC buoys results in RMSE=0.41m. The derived state parameters and imagette information (geo-location, time, ID, orbit number, etc.) are stored in netcdf format for convenient use.

Ground truth data

For tuning and validation three independent sources were taken

- WAVEWATCH III (NOAA) with resolution of ½ degree (spatially interpolated on 20 km grid corresponding to S1 WV vignette size)
- CMEMS (COPERNICUS) with spatial resolution of 1/12 degree.
- NDBC buoy records

Both models have 3h output which were temporally interpolated. The buoy records were also interpolated in time and only records available 30 minutes before or after the S1 acquisition were used.

The whole archive of S1 wv for 2016-2021 results into more than 4Mio imagettes (CMEMS data starts at 2016-05) which can be collocated with model data. Only ~5000 imagettes can be collocated with buoys closer than 50km distance for this time interval. Thus, the buoy data were used only for validations.

Generally, for all collocation only imagettes within -60°<LAT<60° were taken in order to avoid ice coverage.

Processing and SAR features used

The processing of SAR data into CCI products contains 3 blocks:

1. Data preparation: reading, calibration, selection sub-scenes (nine subscenes per imagette) and NRCS filtering
2. SAR features estimation and features filtering
3. Sea state parameters calculation using tuned functions (initial and postprocessing step), filtering, control of results and output.

The SAR first-order features consist of 5 groups where the classical CWAVE parameters introduced in Schulz-Stellenfleth et al., 2007, supplemented in Stopa and Mouche, 2017 were further supplemented at DLR with series of new parameters (Pleskachevsky et al., 2021):

1. NRCS and NRCS statistics: mean intensity, normalized variance, skewness, kurtosis etc. (9 features)
2. Geophysical: Wind using CMOD algorithm.
3. GLCM (grey level co-occurrence matrix) features: variance, entropy, dissimilarity, etc. (8 features)
4. Spectral features based on integrating of image spectrum for different wavelength domains and noise (16 features).
5. Spectral features based on CWAVE parameters as product of normalized image spectrum and 20 orthonormal functions (20 features) and cut-off by ACF (Auto-Correlation-Function).

In the DLR method, in addition to the standard linear combination of the features, the inverse relationship for first-order parameters was applied.

Initial step: Tuning first guess parameters using linear regression

The linear regression model function for first guess sea state parameters was tuned using the 61 S1 relative orbits found to be collocated with NDBC buoys (Pleskachevsky et all., 2021).

The tuning for linear regression coefficients was performed using the collocated orbits for 2017-2018 (total ~250k imagettes, ~150 imagettes per one orbit, each orbit ~4 times per month). For tuning, the wv imagettes were collocated with both CMEMS and with WW3 model data.

The validations were performed using collocated orbits for 2019 including about 220k collocated samples for both S1 wv1 and S1 wv2 imagettes (Round Robin data set). The buoy collocation results into ~3000 samples for the time interval 2017-2019.

In order to compare the ground truth data a cross-validation for SWH was carried out using CMEMS, WW3 and NDBC buoy data. After the model function was tuned independently two times using CMEMS and WW3 data, the coefficients were swapped: sea state estimated with CMEMS tuned coefficients were validated with WW3 data, and vice versa WW3 tuned coefficients were tested with CMEMS wave height. Additionally, two mixed data sets were created, using both CMEMS and WW3 collocations. One dataset was used for tuning and the other was used for validation. All three resulting SWH (CMEMS, WW3, CMEMS&WW3) were validated against the buoys. Table 5.2 presents the results of these cross-validations: first column shows the data sources for tuning, horizontally are specified the data sources for validation. As can be seen, the optimum of RMSE of ~43cm for buoy comparisons is reached, when only CMEMS model is used for training. This might be a consequence of the better spatial model resolution in comparison to WW3.

After cross-validation (see Table 5.2 below) for significant wave height *SWH*, the CMEMS data were selected for entire trainings, as cross-validation with buoys gave better results than the model function tuned with WW3 data. However, not all sea state parameters are included in the CMEMS data. Thereby, the mean period T_{m0}, first moment wave period T_{m1} and second moment period T_{m2} were tuned and validated with WW3 data.

Table 4.2. Cross-validation of total significant wave height using different data sources executed using the 61 collocated orbits. RMSE in meter is given for wv1/wv2

Tuning data	**Validation data source**		
	CMEMS	**WW3**	**BUOYS**
CMEMS	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
CMEMS&WW3	0.34 / 0.39		0.43 / 0.45

The mean accuracy for wv1 and wv2 of the initial SWH, validated against the model data, reached:

- $RMSE_{cmems\_swh}$=0.35m against CMEMS
- $RMSE_{buoy\_swh}$ =0.44m against NDBC buoys

The whole list of RMSE see Table 5.3.

Table 4.3. RMSE for integrated sea state parameters

Sea State Parameter	Symbol	RMSE (both wv1 and wv2)	Data source
Total significant wave height	swh	0.26 m	CMEMS
Mean wave period	Tm0	0.62 sec	WW3
First moment period	Tm1	0.52 sec	WW3
Second moment period	Tm2	0.45 sec	WW3
Dominant swell wave height	swell_swh_primary	0.46 m	CMEMS
Secondary swell wave height	swell_swh_secondary	0.35 m	CMEMS
Windsea wave height	windwave_swh	0.41 m	CMEMS
Windsea wave period	windwave_period	0.62 sec	CMEMS

Postprocessing step: Machine learning

For improvement of first guess values, a machine learning technique was applied. The support vector machine (SVM) technique was applied with epsilon-SVR (regression) function and radial basis function as kernel-type. Additional features for training were added: precise incidence angle and a flag identifying the satellite (S1A or S1B). For tuning of hyperparameters, the data of all S1 wv acquisitions in the period from 2016-06 to 2017-05 with ~500k samples were used for training and all acquisitions in the period of 2017-06 to 2017-12 with ~300k samples were used for testing. Then, for training the final model all data from 2016-05 to 2017-12 and additional data with SWH>4m from all acquisitions in 2018 were applied (~1.2Mio samples).

The postprocessing applies SVM on all initially processed SAR features, while separate models for wv1 and wv2 are trained. The current pair of models “SVM_model_1.2Mio” was then applied for processing the whole S1 wv archive 2014-12 – 2021-02. The postprocessing can easily be improved in the future by replacing the current models with an improved version (e.g. on the basis of using more training data).

4.3.2 Content of Pleskachevsky SAR product

The stored product includes metadata, integrated sea state parameters and quality flags as specified in Table 4.4.

Table 4.4. Variables in CCI-Seastate S1 vw products

coordinate variables	
time	time (in seconds since 1990-01-01)
lat	latitude
lon	longitude
environmental variables for SAR	
swh	total C band significant wave height
swh_uncertainty	RMSE calculated from CMEMS model colocations
swh_quality	quality of C band SAR significant wave height measurement (0 undefined; 1 bad; 2 acceptable; 3 good)
swh_rejection_flags	consolidated Significant Wave height quality flags
Tm0	C band Mean wave period
Tm1	C band First moment period

Tm2	C band Second moment period
swell_swh_primary	C band Dominant swell wave height
swell_swh_secondary	C band Secondary swell wave height
windwave_swh	C band Windsea wave height
windwave_period	C band Windsea wave period
auxiliary variables	
auxiliary measurements	
sea_ice_fraction	sea ice fraction, as taken from CCI Sea Ice products
topography	
bathymetry	ocean depth, taken from GEBCO version 20150318
distance_to_coast	distance to nearest coast, from NASA/GFSC distance to coastline
model auxiliary data	
wind_speed_model_u	U component of the model wind vector, taken from ERA5 reanalysis
wind_speed_model_v	V component of the model wind vector, taken from ERA5 reanalysis
surface_air_temperature	surface air temperature, taken from ERA5 reanalysis
surface_air_pressure	surface air pressure, taken from ERA5 reanalysis
swh_model	Copernicus ERA5 Reanalysis by ECMWF

4.3.3 Validation of Pleskachevsky SAR product

For validation all data in S1 wv archive from 2018-01 until 2021-02 were used. After colocation with CMEMS more than 3.6Mio samples were available. Separately, the samples for buoy collocated orbits in 2019, which were used in the SAR Round Robin were collected. The reached total accuracy (all sea state conditions) of the method for estimation of SWH corresponds to:

- $RMSE_{CMEMS\_wv1}$ = 0.245 m   BIAS= -0.01cm for wv1
- $RMSE_{CMEMS\_wv2}$ = 0.273 m   BIAS= -0.01cm for wv2

The validation for predefined domains of significant wave height is shown in **Figure 4.2**.

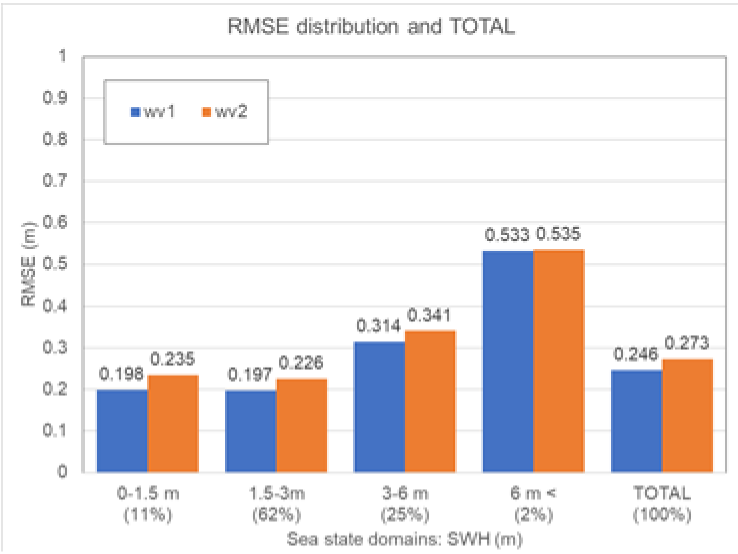

Figure 4.2. Distribution of $RMSE_{CMEMS}$ for different sea state domains.

The monthly estimated total $RMSE_{CMEMS}$ varies around the mean value of 0.24/0.27 and is presented in the following **Figure 4.3**. The monthly RMSE variations are caused by different numbers of storms in different months, since the high waves have a higher RMSE and increase the total RMSE when their percentage grows.

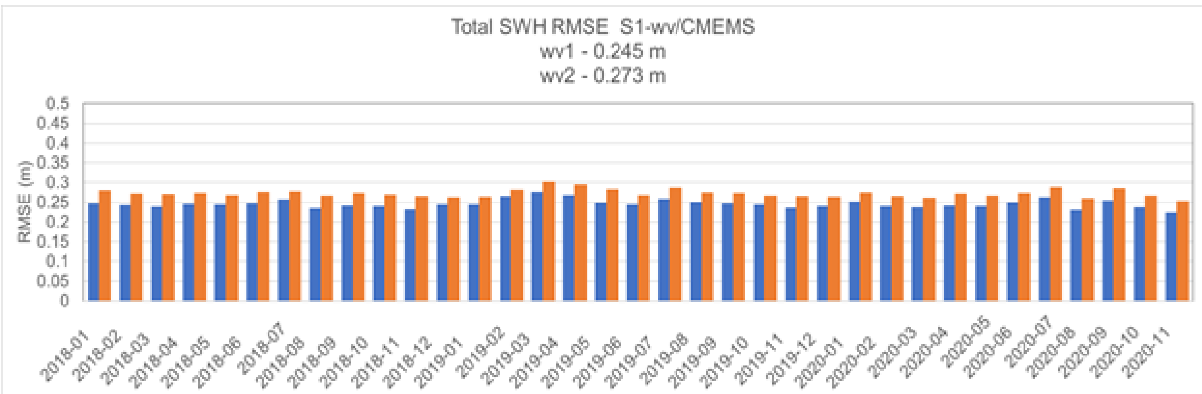

Figure 4.3. Variation of monthly estimated total $RMSE_{CMEMS}$ for wv1 (blue) and wv2 (orange).

A histogram of significant wave height from S1 wv and CMEMS is shown in **Figure 4.4**.

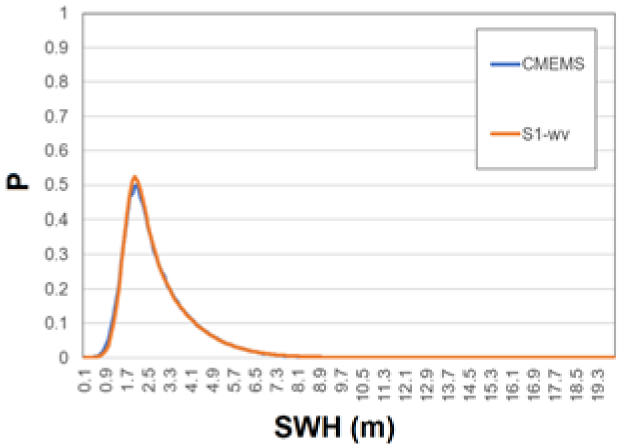

Figure 4.4. Histogram of significant wave heigth SWH from S1 wv and CMEMS, built using 10cm bins for whole S1 wv archive 2018-2021 including all 3.6Mio samples.

The Round Robin (RR) dataset consists of a selection of long collocated orbits with ~10% of all data in 2019 starting close to Antarctic, where a lot of ice is present. The ground truth of the CMEMS model has more uncertainty in this region. Therefore, the validation for the SAR Round Robin dataset results in slight lower RMSE compared to the validation with 3.6Mio samples from 2018-2021.

The distribution of the $RMSE_{CMEMS}$ for different domains of SWH for the RR dataset with 275k CMEMS colocations are presented in Table 4.5:

Table 4.5. RR CMEMS RMSE for *SWH*

SWH domain (m)	Sea State (%)	RMSE wv1 (m)	RMSE wv2 (m)
0.0 - 1.5	~11	0.28008	0.34212
1.5 - 3.0	~62	0.19675	0.22708
3.0 - 6.0	~24	0.30428	0.33291
6.0 <	~2	0.51955	0.55897
Total	100	0.245	0.273

The RR validation against the NDBC buoys results in

- $RMSE_{NDBC\_wv1}$ = 0.415 m BIAS= -0.10 cm for wv1
- $RMSE_{NDBC\_wv2}$ = 0.412 m BIAS= -0.10 cm for wv2

The distribution of the $RMSE_{NDBC}$ for different domains of SWH for the RR dataset with 2000 buoy collocations are presented in Table 4.6:

Table 4.6. RR buoy RMSE for *SWH*

SWH domain (m)	Sea State (%)	RMSE wv1 (m)	RMSE wv2 (m)
0.0 - 1.5	~38	0.37024	0.34758
1.5 - 3.0	~45	0.37357	0.37292
3.0 - 6.0	~17	0.50478	0.50125
6.0 <	~1	0.95620	1.16420
Total	100	0.415	0.412

References:

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

Pleskachevsky, A., Tings, B., Jacobsen, S.: Sea State from Sentinel-1 SAR Wave Mode Imagery for Maritime Situation Awareness, EUSAR 2021, 13th European Conference on Synthetic Aperture Radar, EUSAR 2021, pp.3., IEEE. EUSAR 2021, 29. Mar - 01. Apr 2021, ISBN 978-380075457-1. ISSN 2197-4403. 2021

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

Stopa, J., Mouche, A.: Significant wave heights from Sentinel-1 SAR: Validation and Applications. JGR, Vol., 122, pp. 1827-1848., 2017

4.4 Wave height estimation from Envisat WM using *Li and Huang* methodology (CAS)

4.4.1 Methodology

Over full life cycle of Envisat 2002–2012, the Advanced SAR (ASAR) Wave Mode (WM) data in single look complex (SLC) format were acquired and provided. The WM imagettes, which are dedicated to measurements of ocean waves, are located 100 km apart from each other along the Envisat orbits with alternating incidence angles of 23° or 33°. Each imagette covers a relatively small area of the global oceans approximately with 6 km by 10 km in the two dimensions. As data with HH polarization and data with an incidence angle of 33° are only available during experimental periods, WM data with VV polarization and an incidence angle of 23° were used.

The parametric model "CWAVE_ENV" was developed for ASAR WM data to derive sea state parameters: significant wave height Hs (SWH) and second moment wave period (MWP) (Li et al., 2011). The method applies the widely known CWAVE approach based on linear regression (Schulz-Stellenfleth et al., 2007). The whole Envisat ASAR WM archive from 2002 to 2012 was then processed by Li and Huang (Li and Huang, 2020) for a long-term derivation of SWH and MWP. They further performed a comprehensive calibration and validation for the ASAR derived SWH and MWP using the data of ECMWF buoys and radar altimeters (RA). For each ASAR WM Level1B data a NetCDF formatted product was generated and published.

The original NetCDF format from Li and Huang was finally reprocessed by DLR into CCI format as part of this project. Only the version of the parameters SWH and MWP, which were calibrated according to ECMWF buoy data using the RMA regression method, are used in the CCI products, due to their increased level of accuracy.

4.4.2 Content of Li and Huang SAR product

The list of variables with descriptions stored in NetCDF product for CCI project are listed in Table 4.7.

Table 4.7 List of variables and the descriptions in the CCI ASAR WM sea state NetCDF product.

coordinate variables	
time	time (in seconds since 1990-01-01)
lat	latitude
lon	longitude
environmental variables for SAR	
swh	total C band significant wave height derived using CWAVE_ENV method, calibrated according to ECMWF buoy data using the RMA regression method
Tm2	C band Second moment period Tm02 using CWAVE_ENV method, calibrated according to ECMWF buoy data using the RMA regression method
swh_uncertainty	best estimate of significant wave height standard error

swh_quality	quality of C band SAR significant wave height measurement (0 undefined; 1 bad; 2 acceptable; 3 good)
swh_rejection_flags	consolidated Significant Wave height quality flags 0 for “pass” i.e. acceptable record 1 for a “bad_record” 2 for “land” 3 for “inhomogeneous ASAR imagettes” 4 for “ASAR imagettes in HH polarization” 5 for “ASAR imagettes with an incidence angle not equal to 23°” 6 for “ASAR imagettes in polar regions, i.e. beyond 70°N or 65°S”
auxiliary variables	
heading	satellite heading direction in degrees clockwise from North
angle_of_incidence	Local incidence angle of ASAR imagette center
homogeneity	Homogeneity of ASAR imagettes
normalized_variance	Normalized variance of the pixel’s digital numbers of the ASAR imagette
auxiliary measurements	
sea_ice_fraction	sea ice fraction, as taken from CCI Sea Ice products
topography	
bathymetry	ocean depth, taken from GEBCO version 20150318
distance_to_coast	distance to nearest coast, from NASA/GFSC distance to coastline
model auxiliary data	

wind_speed_model_u	U component of the model wind vector, taken from ERA5 reanalysis
wind_speed_model_v	V component of the model wind vector, taken from ERA5 reanalysis
surface_air_temperature	surface air temperature, taken from ERA5 reanalysis
surface_air_pressure	surface air pressure, taken from ERA5 reanalysis
swh_model	Copernicus ERA5 Reanalysis by ECMWF

4.4.3 Validation of Li and Huang SAR product

Validation with Radar altimeter RA wave data.

The calibrated Ku-band SWH measurements of Jason-1 were used to perform a cross-validation with the calibrated ASAR-derived SWH. The Jason-1 mission provided wave data from December 2001 until July 2013, which covers the lifetime of the ASAR instrument. The data used for validation was provided in the context of the GlobWave project (http://globwave.ifremer.fr/), where the original Jason-1 measurements were reprocessed and quality control flags and calibrated SWH measurements were provided.

In Total, ~23k valid collocated data pairs of Jason-1 and ASAR WM data were obtained by Li and Huang. ASAR SWH shows good consistency with the Jason-1 SWH, with BIAS=0.04 m, total RMSE=0.48 m, correlation coefficient of 0.93 and SI=16.84%.

Validation with in-situ buoy data.

In situ buoy measurements of sea state parameters were used to validate and calibrate the retrieved SWH and MWP based on ASAR WM data. Li and Huang recognized a comprehensive collection of buoy data (649 buoys collected between 2002 and 2012) being provided by the European Center for Medium-Range Weather Forecasts (ECMWF), including data from various networks e.g., NDBC, the Marine Environmental Data Section (MEDS), the Coastal Data Information Program (CDIP) and others. Therefore, the ECMWF-provided buoy data (hereafter referred to as “ECMWF buoy data”) was selected for the evaluation and calibration of the ASAR-derived SWH.

From the resulting comparison (see Table 4.8) of ASAR-derived SWH (excluding the detected outliers of 28k colocations) with the ECMWF buoy SWH a RMSE=0.51 m and a BIAS=0.00 m was estimated. The comparison for ASAR MWP and NDBC buoys shows RMSE=0.65 s and BIAS=0.00 s on the basis of 15k collocations.

Table 4.8. Validation of total RMSE and BIAS for SWH and MWP against RA and buoys.

Sea state Parameter	Data source	RMSE		BIAS	
		raw	calibrated	raw	calibrated
SWH	Jason-1 RA	0.48 m	0.53 m	0.04 m	0.18 m
	ECMWF buoys	0.51 m	0.51 m	-0.06 m	0.00 m
MWP ()	ECMWF buoys	0.67 s	0.65 s	-0.19 s	0.00 s

For different sea state conditions, the RMSEs are distributed according to Table 4.9.

Table 4.9. Variations in the bias (Buoy – ASAR) and RMSE for varying sea state conditions

SWH Range (m)	Description	No. collocations	BIAS (m)		RMSE (m)	
			Raw	Calibrated	Raw	Calibrated
0.50 – 1.25	Slight	4548	-0.43	-0.22	0.54	0.44
1.25 – 2.50	Moderate	13031	-0.16	-0.03	0.40	0.41
2.50 – 4.00	Rough	7362	0.14	0.13	0.52	0.59
4.00 – 6.00	Very rough	7362	0.38	0.18	0.74	0.75
6.00 – 9.00	High	507	0.46	-0.04	0.89	0.87
9.00 –14.00	Very high	19	0.74	-0.06	0.92	0.62

References:

Schulz-Stellenfleth, J., König, Th. and Lehner, S. (2007), An empirical approach for the retrieval of integral ocean wave parameters from synthetic aperture radar data. J. Geophysical. Res., Vol. 112, pp. 1-14.

Li, X.-M., Lehner, S. and Bruns, T. (2011), Ocean wave integral parameter measurements using Envisat ASAR wave mode Data. IEEE Trans. Geosci. Remote Sens. 49, 155–174.

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.

6 Annex A: Differences with version 1 dataset

The main differences between CCI Sea State version 1 and version 2 datasets include:

- a **new retracking algorithm** (WHALES) applied consistently to all version 2 altimeter L2P products to retrieve the significant wave height (SWH) from altimeter waveforms, whereas in version 1 we were using and correcting the SWH provided by space agencies which were using different retrackers (see section 4.1.2.1)
- a **more limited coverage** (2002-2019) in version 2 for altimeter products instead of 1992-2018 in version 1; missions previously available in version 1 such as ERS-1, ERS-2, GFO and Topex are no longer available in version 2. This is due to the choice of applying a consistent retracker for all missions; additional effort and investigation are required to properly address these older missions which will be undertaken for a future release. Jason-1 and Jason-2 also have a smaller temporal coverage in version 2 than in version 1 as graveyard orbit phases were excluded from processing. They will be added in a next release.
- a **revised editing and SWH outlier removal scheme** for version 2 altimeter L2P products (see section 4.1.2.4)
- a **revised estimation of cross-mission SWH adjustment corrections** and **SWH uncertainties** for version 2 altimeter L2P products (see section 4.1.2.5 and 4.1.2.6)
- **no more cross-mission adjustment corrections to sigma0** are provided in version 2 altimeter L2P products, this will be addressed in a future release
- some **format and content changes** in version 2 altimeter L2P products:
  - added **model SWH from ERA5** as an ancillary variable (*swh_model*)
  - fixed values in **distance to coast** (distance_to_coast) now in meters (instead of km), as specified by the variable units attribute
  - ancillary **sea level variables are now taken from CMEMS** (SEALEVEL_GLO_PHY_L3_REP_OBSERVATIONS_008_062 dataset) instead of CCI Sea Level, as it is more comprehensive and more frequently updated. As a consequence, the CMEMS variables *mdt* and *sla_unfiltered* are now replacing the former CCI Sea Level *mean_sea_surface*, *sea_state_bias* and *corssh* variables.
- addition of **SAR L2P products** providing SWH and spectral wave parameters for Envisat, Sentinel-1 A and Sentinel-1 B missions (see section 5)

<End of Document>