End-to-end ECV Uncertainty Budget (E3UB)

version 5.2   09 March 2026

Contents

Author	Approved	Signature	Date
Manuel Lopez	Annabelle Ollivier		9 March 2026
ESA Acceptance	Digitally signed by Sarah Connors Date: 2026.03.30 14:34:45 +01'00'		

Issue	Date	Comments
1.0	8 July 2019	First version for ESA approval
1.1	20 Sept. 2019	Minor revisions after comments from ESA
2.0	21 May 2020	Additional detail and recommendations for version 2 dataset
3.0	8 Sept 2021	Additional information on SAR retrievals and error analysis from Climate Assessment Report
4.0	10 October 2024	Update for Phase 2
5.0	15 December 2025	Revised to include discussion of extra products generated by options
5.1	26 february 2026	Updated with options (TSX, CFOSAT, ERS-Gyroless)
5.2	9 March 2026	Update after review (CFOSAT, ERS-Gyroless)

List of Acronyms

ADP	Algorithm Development Plan
AoI	Area of INterest
ASR	Algorithm Selection Report
ATBD	Algorithm Theoretical Basis Document
CAR	Climate Assessment Report
cci	Climate Change Initiative
CMEMS	Copernicus Marine Environment Monitoring Service
DD	Delay-Doppler
DtC	Distance to Coast
E3UB	End-to-End ECV Uncertainty Budget
ECMWF	European Centre for Medium-range Weather Forecasts
ECV	Essential Climate Variable
FFT	Fast Fourier Transform
GDP	Global Drifter Programme
GDR	Geophysical Data Record
GPS	Global Positioning System
GTS	Global Telecommunication System
IFS	Integrated Forecasting System
L4	Level 4
LRM	Low Rate Measurement
LUT	Look-Up Table
MAD	Median Absolute Deviation
MERIS	Medium Resolution Imaging Spectrometer
MLE	Maximum Likelihood Estimator
MWR	Microwave Radiometer
NRCS	Normalized Radar Cross-Section
NWP	Numerical Weather Prediction
OSTST	Ocean Surface Topography Science Team
PHCP	Percentage of High Correlation
PLRM	Pseudo Low Rate Measurement
PTR	Point Target Response
ASR	Algorithm Selection Report
PVP	Product Validation Plan
RA	Radar Altimeters
RR	Round Robin
R.m.s.	Root mean square
RMSE	Root mean square error
S3A	Sentinel-3A
S3B	Sentinel-3B
SAR	Synthetic Aperture Radar
SI	Scatter Index
SSH	Sea Surface Height
SWH	Significant Wave Height
S.D.	Standard Deviation
TCT	Triple Collocation Technique

WHALES	Wave Height Adaptive Leading Edge Subwaveform (retracker)
w.r.t.	with respect to
WV	Wave (mode for SAR)

1. Introduction

This document presents version four of the End-to-end ECV Uncertainty Budget (E3UB) for **Sea_State_cci**, deliverable 2.3.V2 of Phase II of the project.

The calculation of significant wave height (Hs) from altimetry data is a relatively straightforward inversion of a direct measure of the height distribution of reflecting facets. Unlike derivation of sea surface height, the estimate is not affected by atmospheric or ionospheric conditions, and is not dependent upon models of tides or atmospheric pressure or liquid water content. All retracking algorithms fit a shape to the observed waveform, and all the LRM ones use the slope of the leading edge to infer the breadth of the distribution of reflecting facets.

Nevertheless there will be some errors in the inversion, due to the effects of noise, and of incorrect assumptions in the model. There are also errors associated with the production of gridded products. This version of the End-to-End ECV Uncertainty Budget (E3UB) summarises the state of the knowledge two years into the second phase of the *Sea State CCI* project. The major change from the previous version is the inclusion of discussion of uncertainty in the extra products generated by the CCI options.

Illustrations are generated for January 2017 (or else Dec. 2016-Feb 2017) as this is boreal winter and is expected to have some large wave height values, and it is a period when Jason-2 and Jason-3 are both operating well, but not on the same tracks (so providing independent spatial sampling).

This version of E3UB also covers assessment of the two SAR algorithms (Ifremer/Stopa and DLR) applied to Sentinel-1 data, noting evaluations separately for incidence angles of 23° and 36°. There is also a brief discussion of the derived trends in gridded composite products and the levels of uncertainty associated with them.

2. Errors in the Instruments / Algorithms

2.1 Altimeter (LRM)

There are two types of errors associated with the algorithms — those due to inadequacies of the inversion model, which will be consistent for the same conditions and can lead to an under- or over-estimation bias, and those that are caused by the sensitivity of the model to instrumental noise in the altimeter.

2.1.1 Instrument Noise

The principal cause of high-frequency variability is "fading noise" (or multiplicative noise) due to the signal at each waveform bin being the sum of many independent contributions with random phase. Each realization of fading noise is usually taken to be independent from all its neighbours, and should thus cause errors that are independent between successive high-rate estimates. This component of error can be estimated by applying a low-pass filter to determine the underlying geophysical variation, and examining deviations from that, or, more simply, by utilising σ_{Hs} (the S.D. of the high-frequency estimates contributing to each 1 Hz record). This assumes there are no significant true variations in the geophysical signal on such small scales. This will generally be true away from the coast and away from fronts, so by examining statistics on a global basis these effects are minimised. Figure 1a shows σ_{Hs} as a function of Hs for Jason-3.

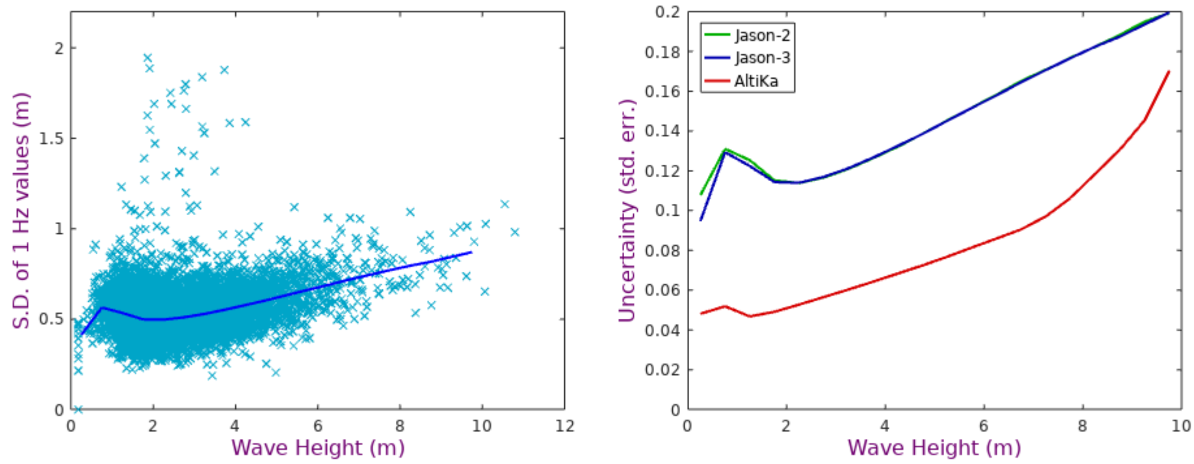

Figure 1 : a) Left panel shows a scatter plot of σ_{Hs} as a function of Hs for Jason-3's MLE-4-based estimates during Jan. 2017. The cyan crosses only show a small subset of the points, but the blue line shows the mean relationship derived from all data (averaged in 0.5 m bins). b) Right panel shows the effective standard error of the 1 Hz mean values, by dividing the curves for Jason-2 and Jason-3 by √19 and that for AltiKa by √39.

The analysis in Fig. 1a shows the results for the standard MLE-4 algorithm in the Jason-3 GDRs. The mean value of σ_{Hs} is ~0.5 m at low wave heights and increases with Hs. There is also a slight increase at values of around 0.5 m because the leading edge is poorly resolved in such wave conditions (only bracketted by 2 or 3 wavebins) and thus the slope is hard to estimate accurately. A reduction again in σ_{Hs} for Hs<0.3m is not trustworthy, as it simply reflects that all derived negative values are set to zero, thus reducing the variability. The magnitude and shape of the curve are properties of the altimeter and the retracker applied. Some of the retrackers being developed within the *Sea State cci* project yield lower variability (Fig. 2). These plots show how the 1 Hz variability ("Noise") is very different for different retrackers applied to the same data (Jason-3 and Sentinel-3A in this case), with, in most cases, the variability increasing with wave height. However, some of the methodologies encompass a degree of along-track smoothing, and thus this measure of noise is not always a good estimate of the variability caused by the fading noise. Note that the analysis shown in Fig. 2b encompasses algorithms applied to both SAR and PLRM mode data from Sentinel-3A.

Significantly different results are found for AltiKa (the only altimeter to operate at Ka-band to-date). Its higher operating radar frequency permits useful operation with a higher number of independent pulses per second. Secondly, it has a narrower emitted pulse (smaller PTR width) and narrower wavebins within the waveform, which ensure that there is better sampling of the leading edge at low wave height conditions. Therefore the slight rise in σ_{Hs} as Hs approaches zero is less pronounced.

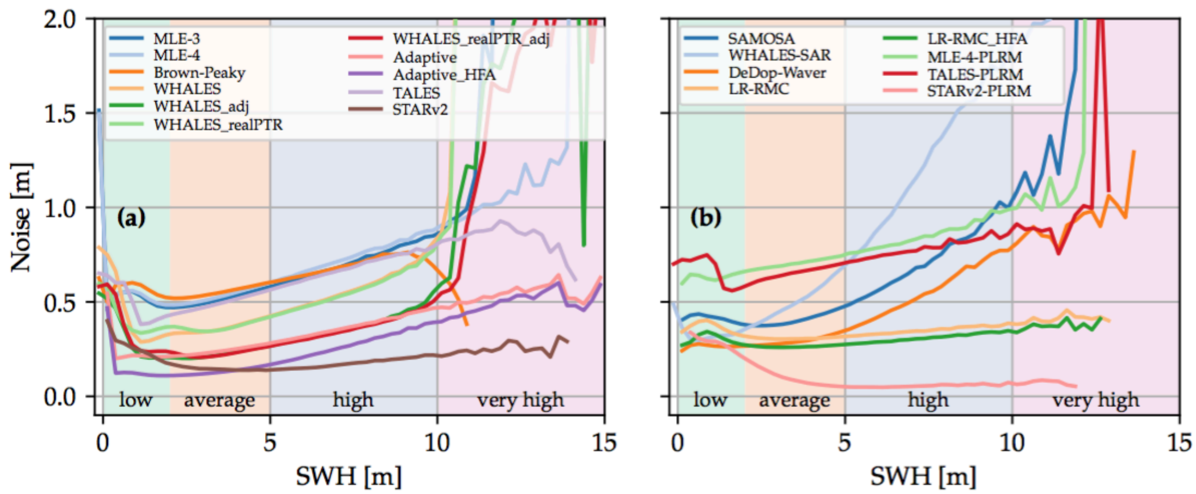

Figure 2 : Noise level of the individual retrackers as a function of significant wave height (SWH) for a) J3- and b) S3A-retracking algorithms with the sea state noted at the bottom. The estimates of noise for "very high sea states" may not be reliable due to limited no. of observations and possible association with intense rain events. Also, the data at apparently very high sea state may not be reliable because they could be outliers that are not caught by the quality flag, which often generates high (although realistic) SWH values [Taken from Schlembach et al. (2020)].

2.1.2 Algorithm Bias

Error in the assumptions used for the inversion algorithms can lead to biases that will not average out over many independent waveforms. Most LRM algorithms model the slope of the leading edge as being due to the combined effect of emitted pulse (PTR) and the smearing due to reflecting from surface facets at different heights. This is often expressed as a composite width, σ_C:

$$\sigma_C^2 = \sigma_P^2 + (Hs/2c)^2 \qquad (1)$$

where σ_P is the width of the Gaussian modelling the Point Target Response (PTR). Systematic errors can then be introduced by uncertainty in the appropriate value for σ_P (especially if long-term space exposure is believed to have changed the value from that recorded during on-ground testing) or when the PTR cannot be reliably modelled by a Gaussian curve. Given that the Hs term will dominate σ_C for Hs>2m, the concerns about the actual shape and width of the PTR are only pertinent for low Hs conditions. Figure 3a shows the inferred instrument correction on Jason-3 to compensate for the real PTR shape.

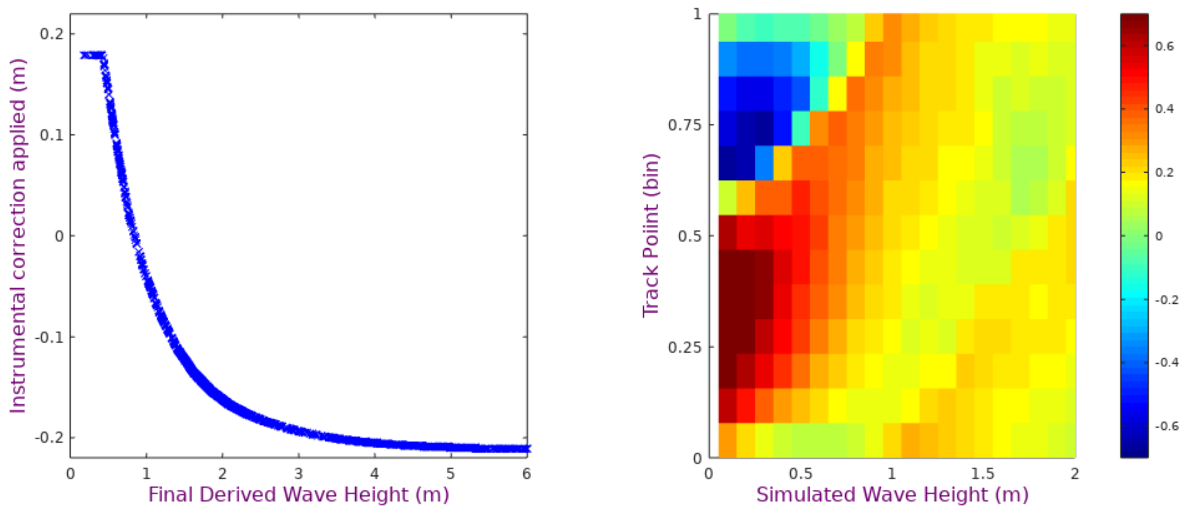

Figure 3 : a) Left panel shows LUT correction for MLE-4 algorithm applied to Jason-3. b) Right panel shows mean bias observed using simulation results for the WHALES algorithm, with the bias varying according to both wave height and the position of the leading edge.

CLS have developed a numerical retracker which uses measured PTR shape in the inversion rather than the Gaussian approximation; others have applied a Look-Up Table (LUT), based on the values shown in Fig. 3 to produce a correction after the inversion. The correction is significant, but once compensation has been made for the real PTR it is less clear how large the error is. Simulation work performed within the *Sea State CCI* suggests that there may still be large errors after correction because the effect depends upon the position of the waveform within the window. Although averaging over many successive waveforms will reduce this effect due to positioning, it may not disappear because of a systematic bias caused by the on-board tracker placing the reception window differently when approaching and receding from the Earth (i.e. principally when heading equator-ward or poleward). However such effects only seem noticeable for Hs significantly below 1m. An attempt to code the variation in PTR effect with position of leading edge within the tracker window, did not achieve the goal of reducing errors (see discussion of WHALES_realPTR in Schlembach et al. (2020)).

2.2 Altimeter (DD)

The waveform shape associated with SAR altimetry is different from that for LRM, with changes in wave height affecting both the leading edge and to a lesser extent the trailing edge of the shape. Thus a greater number of waveform bins show sensitivity to Hs, which may be expected to improve the resilience of estimates to the effect of fading noise. However with the current default algorithm on Sentinel-3A, the variability of the 20 SAR estimates in a second (Fig. 4a) is similar to that for LRM. However, significant advances have been achieved in the past year, with algorithms developed by CLS/CNES, TUM and isardSAT (LR-RMC, WHALES-SAR and DeDop-Waver respectively) which have reduced noise levels compared with SAMOSA (see Fig. 2b).

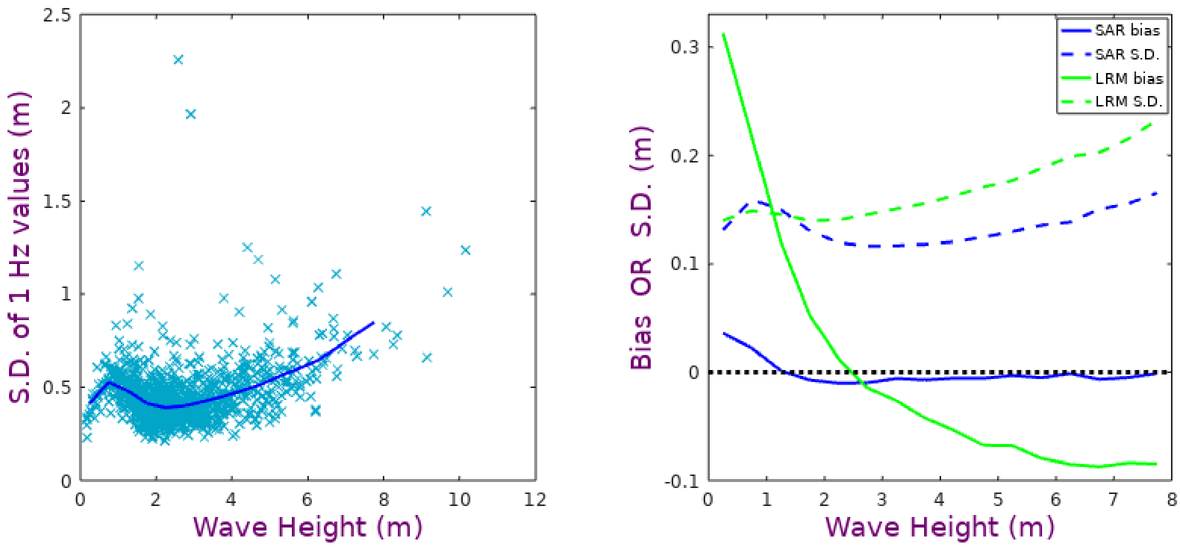

Figure 4 : a) Left panel shows σ_{Hs} (S.D. of values in a 1-second interval) as a function of mean Hs. Data are from Sentinel-3A cycle 033. Cyan crosses show a subset of the points, the blue line shows the mean relationship derived from all the data in the cycle. b) A comparison of S3B data with near-simultaneous S3A data during S3A cycle 033, with S3B first in LRM mode, then later in SAR mode, whilst S3A is always in SAR mode. The 'bias' shows the mean of S3B-S3A and the S.D. shows the variability about this mean.

At present there is a disparity between the Hs retrievals in LRM and SAR mode. When both Sentinel-3 altimeters are in SAR mode there is minimal bias between them and a S.D. of 0.12 to 0.15 m (Fig. 4b); however with S3B in LRM mode its bias relative to S3A varies between 0.3 m at low Hs to almost -0.1m at high Hs, with greater variability of the difference. Part of this difference in behaviour is due to the narrow footprint that can be achieved with the SAR processing: when there is very long wavelength swell it is possible that the footprint does not contain the full variation in height of reflecting facets (Fig. 5a). This depends upon the direction of propagation of the swell relative to the flight direction of the satellite, but analyses using swell direction inferred by models appear to quantify this effect (Fig. 5b).

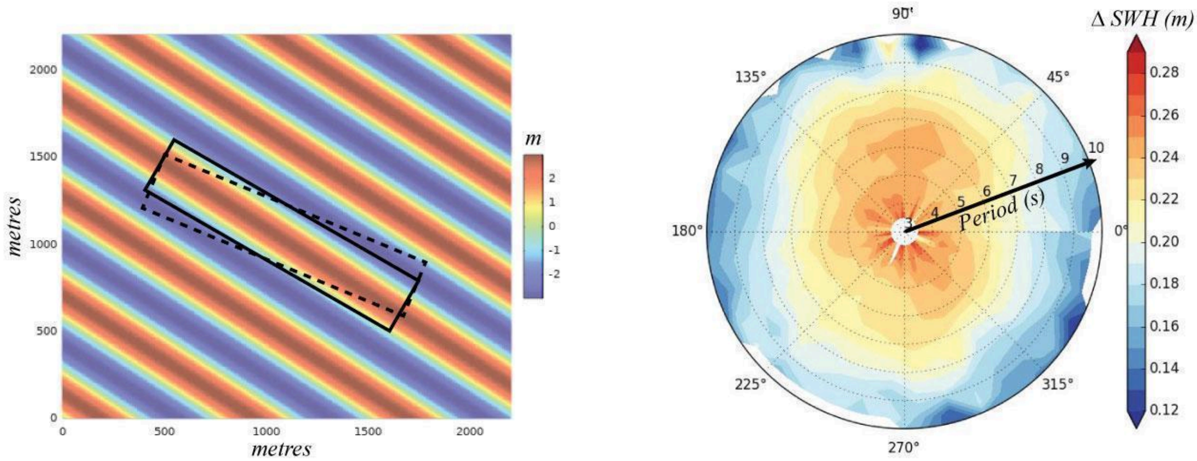

Figure 5 : a) Left panel shows schematic of a simple long wavelength swell field (colours indicating the height of the water surface), with the narrow SAR mode footprint almost aligned with direction of swell. b) Difference between SAR and PLRM estimates of wave height as a function of angle between modelled swell and altimeter flight direction. [Illustration taken from presentation by M. Raynal (CLS).] Observations are for wave heights in the range 2 to 3 m. The relative bias is independent of direction for short period waves (centre of diagram), whereas for the longest periods (and wavelengths) the bias is altered most rapidly if the footprint width and the wave crests are aligned (0˚ and 180˚ in the diagram).

This is an area of intense investigation, and will change according to which version of the Processing Baseline is used. Figure 6 shows a comparison of maps of the mean wave height averaged over the whole of 2019 for Sentinel-3A (PLRM), Sentinel-3A (SAR) and Jason-3 (LRM), with the S3A data being from the recent (Jan. 2020) reprocessing. There is a clear regional pattern in the difference of Sentinel-3A SAR and PLRM estimates that appears to be linked to the mean Hs value; however this does not imply that the bias for 1 Hz estimates is simply a linear function of Hs, as the annual mean values may be affected by the frequency with which one or other algorithm returns an anomalous (e.g.near zero) value but the data pass flagging checks. This can be noted in the histograms (Fig. 6d) and emphasises the importance of assessing data flagging along with accuracy of estimates.

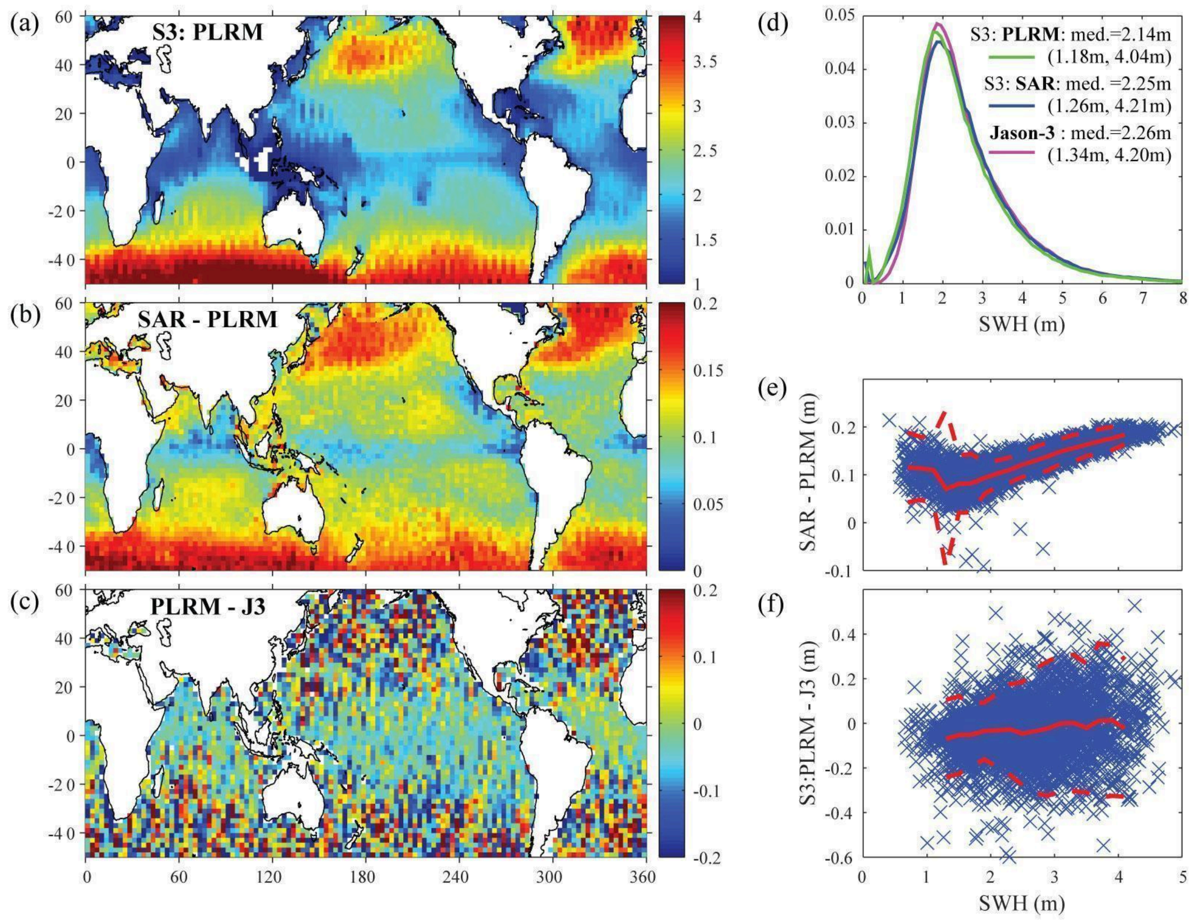

Figure 6 : Comparison of altimeter SWH data for 2019. a) Average of all 2019 for Sentinel-3 PLRM data b) Mean bias of Sentinel-3 SAR mode data to those from PLRM c) Difference of mean fields from S3:PLRM and Jason-3 (LRM) d) Probability distribution functions of SWH data for Jan. 2019. Numerical values are the median and (in brackets) the 10th and 90th percentiles. Data are limited to 50˚ S to 60˚ N to avoid effects of undetected sea-ice. Comparison of grid box means for 2019 for e) S3:SAR relative to PLRM and f) PLRM relative to Jason-3. Full red lines show mean relationship in 0.2 m wide bins with the dashed lines showing ±2 std. dev. For these last 2 plots some data exist outside the axes shown, but these are relatively few. [Taken from Quartly et al. (2020)].

The great disparity between LRM and DDA retrievals of Hs, and the fact that the DDA measurements would not be *additional* but possible *replacements* for LRM ones, has led the project to conclude that only LRM algorithms will be used for all the altimeters.

2.3 Synthetic Aperture Radar

2.3.1 Sentinel-1 WV

The SAR methods were developed and validated for Sentinel-1 Wave-Mode WV data. WV vignettes (imagettes) 20 km by 20 km were acquired every 100 km along the orbit with two different incidence angles approximately 23° (wv1, imagettes with odd numbers 1,3,5..) and 36° (imagettes with even numbers 2,4,6…..) respectively. Vignettes on the same incidence angle are separated by 200 km.

The estimation of sea state integrated parameters (significant wave height, periods Tm1, Tm2, etc.) from SAR data is based on direct processing of intermediate parameters (e.g. variance, spectral parameters, etc.) from NRCS measured on image subscenes. The estimating is rarely affected by atmospheric conditions. In most of the cases, their influence is taken into account by algorithms. In the DLRI method, around 1% of data are discarded (flagged) due to atmospheric and other effects. For example, under strong winds, the sea surface can be completely disrupted by wave breaking and also man-made artefacts like ships, ship wakes or wind farms can spoil the sea state signal.

When the algorithms for SAR sea state retrieval are tuned by utilizing models and buoys as ground truth, then the uncertainties of both ground truths are propagated into the resulting algorithm's accuracy. Worldwide, model data (e.g. WWIII, CMEMS) and buoys agree with an RMSE of ~0.25 m for general conditions which, however, can reach an RMSE of ~0.5 m under storm conditions, mostly due to a relatively small temporary shift of the storm peak passing through a buoy position.

The comparisons show, both algorithms DLR and Ifremer (Stopa) have near identical SWH accuracy. DLR has a little better RMSE, Ifremer has slightly fewer invalid data (No-Sea-State percentage). So, by CMEMS validation, the total RMSE averaged for wv1 and wv2 is 0.263 m (DLR) and 0.273 m (Ifremer), the no-sea-state percentage is 1.08 % (DLR) and 0.24 % (Ifremer). For the NDBC validation similar numbers are obtained: RMSE of 0.414 m (DLR) and 0.445 m (Ifremer) and no-sea-state percentage of 1.45 % (DLR and 0.83 % (Ifremer).

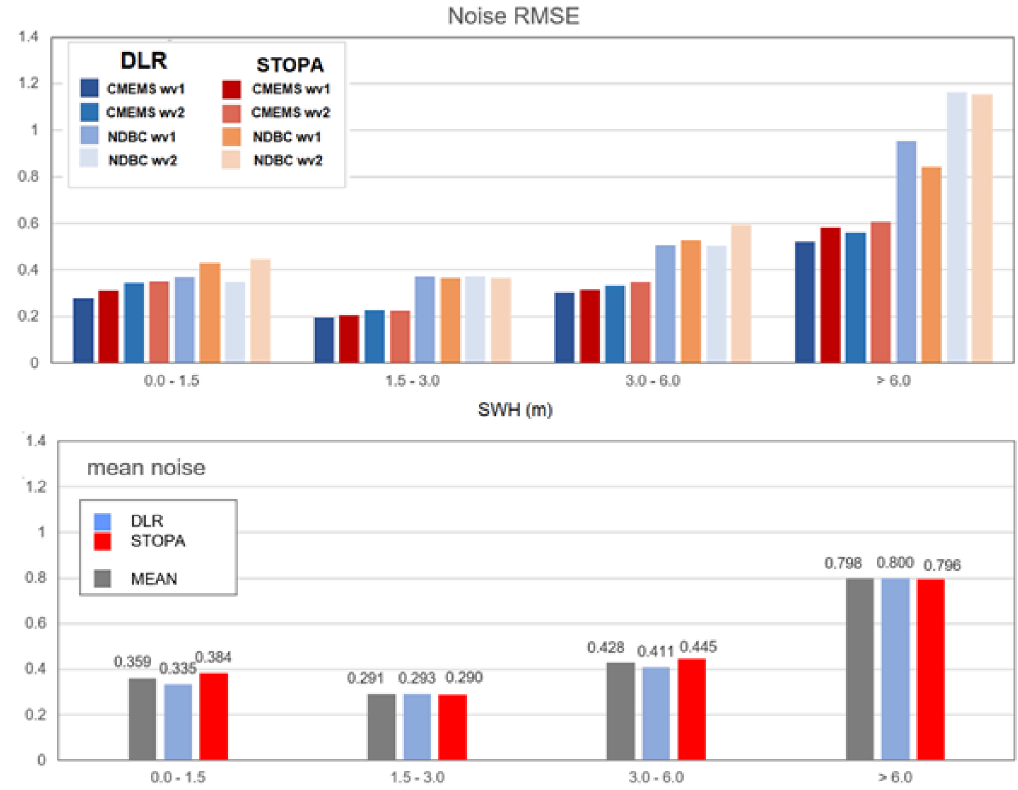

Figure 7 : RMSE distribution for four sea state domains and mean RMSE noise estimated. The mean values in the second line were built with weighting of 0.5 for both CMEMS and NDBC.

Figure 7 shows the RMSE distribution for four considered sea state domains and mean RMSE noise. In the first line all ground truth comparisons are shown. DLR RMSE values are marked with blue-palette colors, the Ifremer RMSE values are marked with red-palette colors. In the lower graph, for each sea state domain, the mean values are built with weighting of 0.5 for both CMEMS and NDBC RMSE's. The grey colored bars present the mean value for both DLR and Ifremer algorithms. Note, for sea state>6m the NDBC has only 30 collocations for buoys in 50 km radius from S1 WV imagette borders, while CMEMS has around 5000 exact geo-located collocations.

The noise analysis shows similar trends for both algorithms:

- the accuracy averaged for forecast model and buoy measurements is of the order of ~0.3 m for the sea state category of moderate sea state, where ~60% of all data points are located
- this averaged accuracy declines with increasing SWH values to ~0.42 m for rough sea state with SWH about 4-5m and to ~0.8 m for very high sea state over 6m SWH. This is connected to both accuracy of the SAR methods and to an increased error in the ground truth, which also grows proportional to the larger sea state values. However, if one connects the local RMSE with the mean value for this sea state domain, the same scatter index SI can be obtained.
- for low sea state (SWH<1.5 m) some difficulties can be seen in comparison with moderate sea state, the lower accuracy does not match to the tendency of higher error with higher sea state described before. This effect is connected to the specifics

of SAR imaging of sea surface, where the short and small waves cannot be imaged individually, but are only visible as noise. Although it is possible to derive their characteristics from the noise, the accuracy is slightly lower in contrast to visible wave patterns.

Note, the validation was carried out for -60°<LAT<60° in order to avoid ice coverage. However, as analyses showed, sea ice can be encountered within -60°<LAT<-55° so that both SWH, from the model and estimated from S1 WV, may be affected. For the DLR algorithm a number of outliers and no-sea-state flags are especially high in this area. If the validation area is reduced to -55°<LAT<60°, the resulting total RMSE improves by ~1.5 cm.

Fig. 8 shows the distribution of squared difference between SWH estimated form S1 and ground truth (CMEMS). On the graph all 270.000 round robin data points are shown, the horizontal axis means latitude of acquired and compared data. As can be seen, in both algorithms, a jump of this differences appears at LAT<-58°. This inconsistency is due to sea ice.

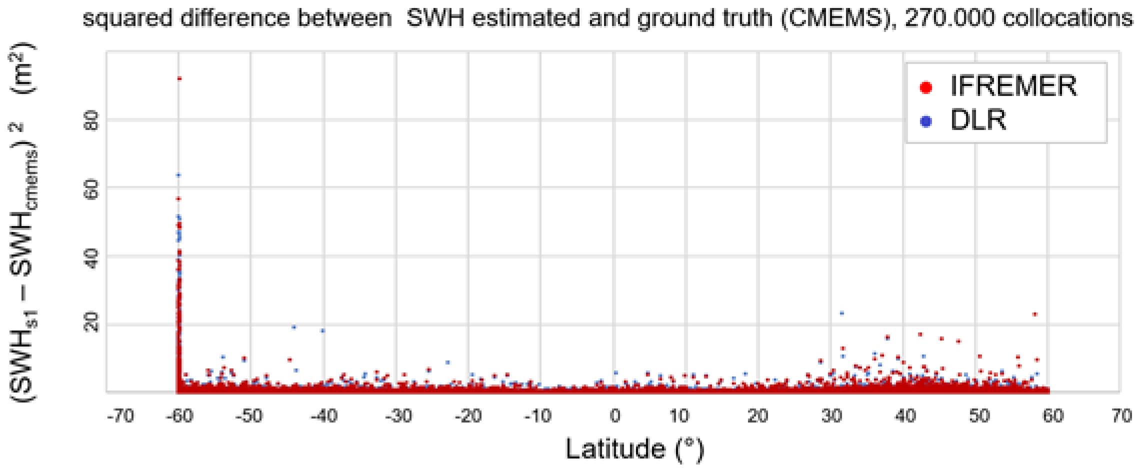

Figure 8 : Distribution of squared difference between SWH estimated form S1 and ground truth (CMEMS). The analysis was restricted to -60°<LAT<60° in order to minimise effects of ice coverage; however, for both algorithms, the effect of sea-ice is quite visible for LAT<-58°. As can be seen this masking is not enough to completely eliminate the ice uncertainty produced by both: CMEMS and S1.

In order to prove that the method does not include any smoothing, along-track comparisons and comparisons of PDFs were carried out. A typical example of an along-orbit comparison for a long overflight of around 12,000 km is presented in Fig. 9.

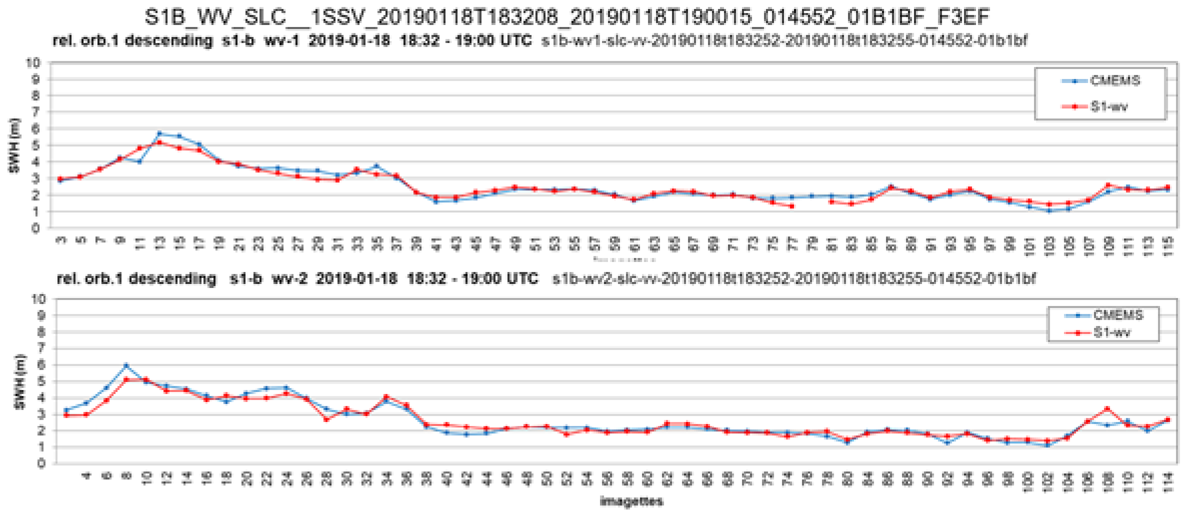

Figure 9 : An example of an along-orbit comparison for one long overflight of around 12,000 km (product ID S1B_WV_SLC__1SSV_20190118T183208_ 20190118T190015_014552_01B1BF_F3EF) plotted separately for wv1 (odd number imagettes on the first graph, 200 km between each points) and wv2 (even number imagettes on the second graph) vignettes. Red points mean DLR SWH and blue are CMEMS SWH temporally interpolated 3h model outputs.

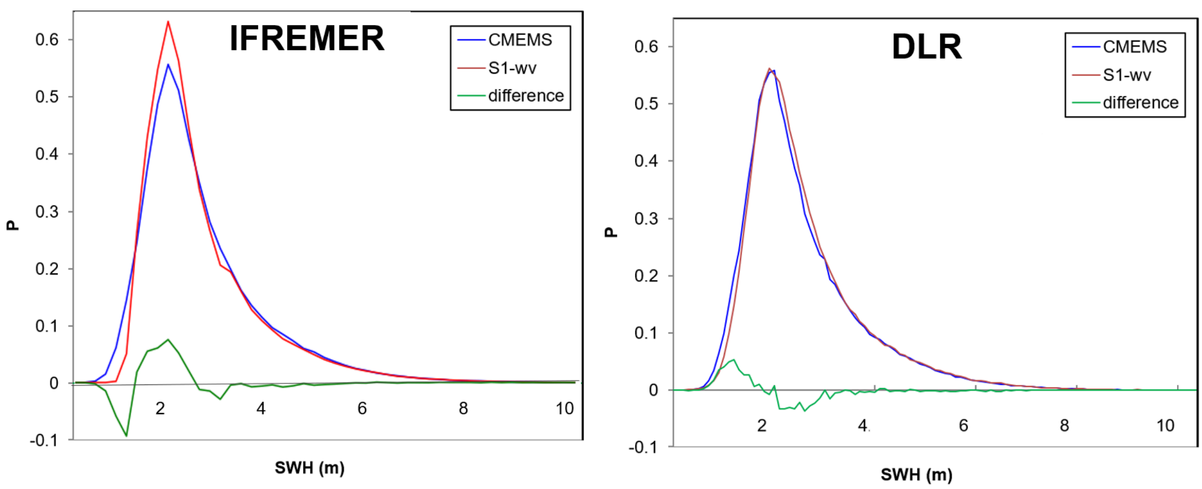

Figure 10 : P.d.f. of SWH distribution for CMEMS and estimated from S1 WV imagettes. a) Ifremer (Stopa) algorithm, b) DLR.

Figure 10 shows the p.d.f. of the SWH distribution in ground truth (CMEMS, blue curves in both graphs) and of the SWH estimated from S1 WV vignettes (red curves) constructed from both algorithms. The Ifremer algorithm comparison is in the left-hand panel, the DLR results are on the right, the differences between PDF-S1 and PDF-CMEMS for both algorithms are green. As the DLR algorithm was tuned with CMEMS data, the comparison for that algorithms leads to slightly lower differences. The Ifremer algorithm was tuned using altimeter data; this probably results in a slight smoothing effect for low wave heights.

The bias comparisons are shown in Fig. 11. In the top graph, the original values for both algorithms are presented for each sea state bin. The values scatter strongly with “-/+” signs, so the centre graph shows the ABS(BIAS). The figure’s bottom graph presents averaged values for CMEMS (green) and NDBC (magenta). The strongest BIAS can be seen for SWH>6m for NDBC. However, this value is based on 30 NDBC collocations, while the CMEMS has around 5000 collocations for SWH>6m.

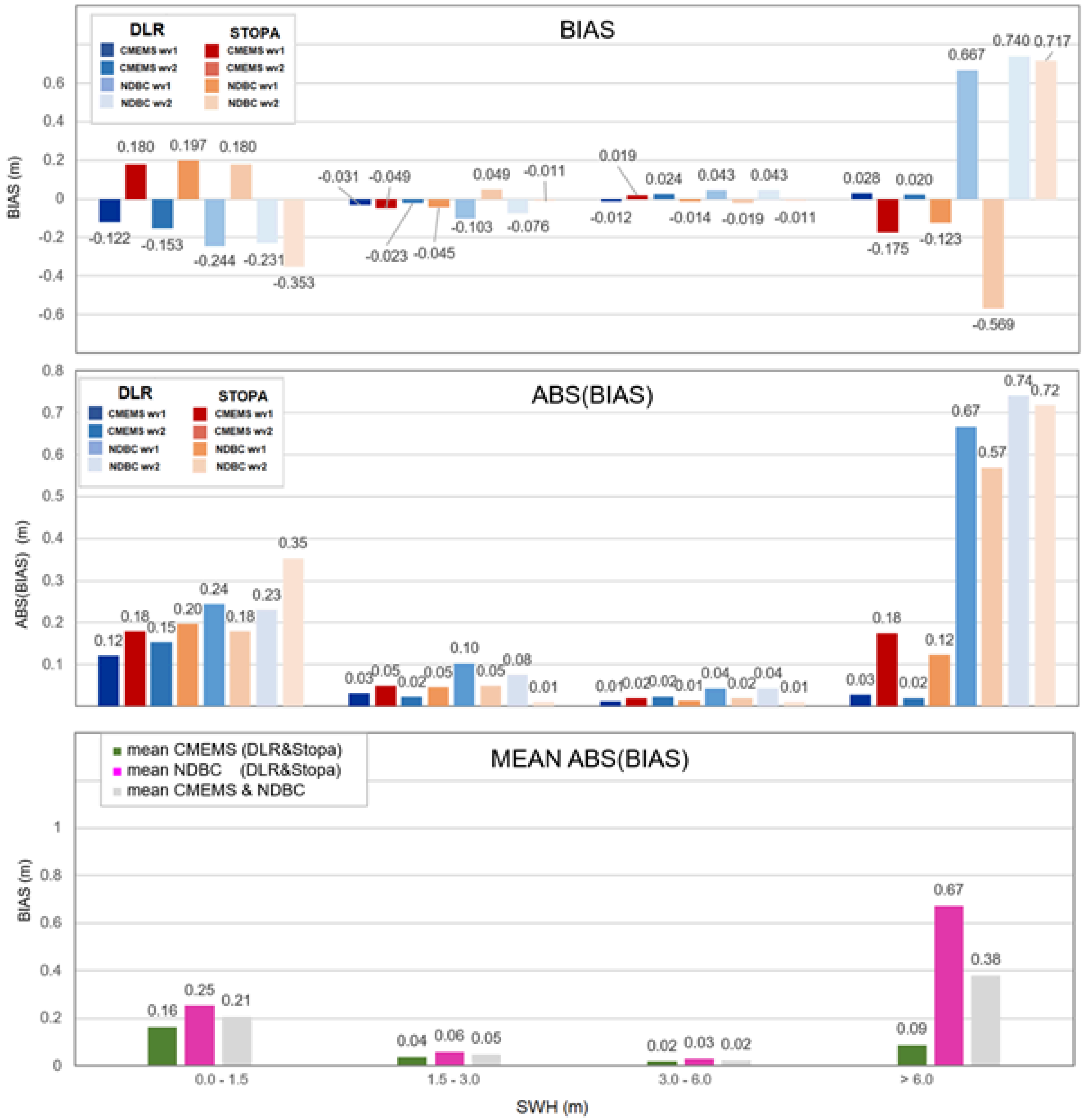

Figure 11 : Bias of both DLR and Ifremer (Stopa) algorithms (top graph), ABS(BIAS) (centre graph) and mean for CMEMS and NDBC for each sea state domain (bottom graph). Grey bars show averaged ABS(BIAS) for CMEMS and NDBC with weighting of 0.5.

2.3.2 Sentinel-1 IW and EW

The validations are based on processing of the whole ESA archive of S1 IW and EW and partially published in Pleskachevsky et.al., 2024. Table 1 shows averaged values for RMSE against MFWAM (CMEMS) model conducted for all data 2020 for two Areas of INterest (AoI): (1) West coast of Canada/USA including Aleutian Islands, and (2) Atlantic and European waters including Mediterranean, North, Baltic and Black seas (see Fig. 12). The scatter in Figs. 14-16 shows comparisons for all worldwide scenes without exclusion acquired in January 2020 (~15000 IDs, due to 5 km processing raster ~10M samples/month) for S1-IW and for all months in 2020 for S1 EW (~6000 ID/moths, due to 17.7 km processing raster ~3M samples per year, all validations were carried out -50°<LAT<60° to avoid ice coverage).

Table 1 : total RMSE for integrated sea state parameters (based on data Pleskachevsky et al., 2024)

Parameter	Abb.	Unit	Satellite mode		
			S1 IW	S1 EW	S1 WV (wv1/wv2)
total significant wave height	swh	m	0.42	0.51	0.24 / 0.28
mean wave period Tm0-1	Tm0	s	0.88	0.92	0.46 / 0.51
first moment wave period	Tm1	s	0.97	0.85	0.51 / 0.56
second moment wave period	Tm2	s	0.96	0.86	0.46 / 0.51
wave height swell dominant system	swell_swh_primary	m	0.57	0.60	0.42 / 0.47
wave height swell secondary system	swell_swh_secondary	m	0.38	0.44	0.41 / 0.46
significant wave height windsea	windwave_swh	m	0.48	0.57	0.40 / 0.46
mean period windsea	windwave_period	s	0.97	0.95	0.62 / 0.67

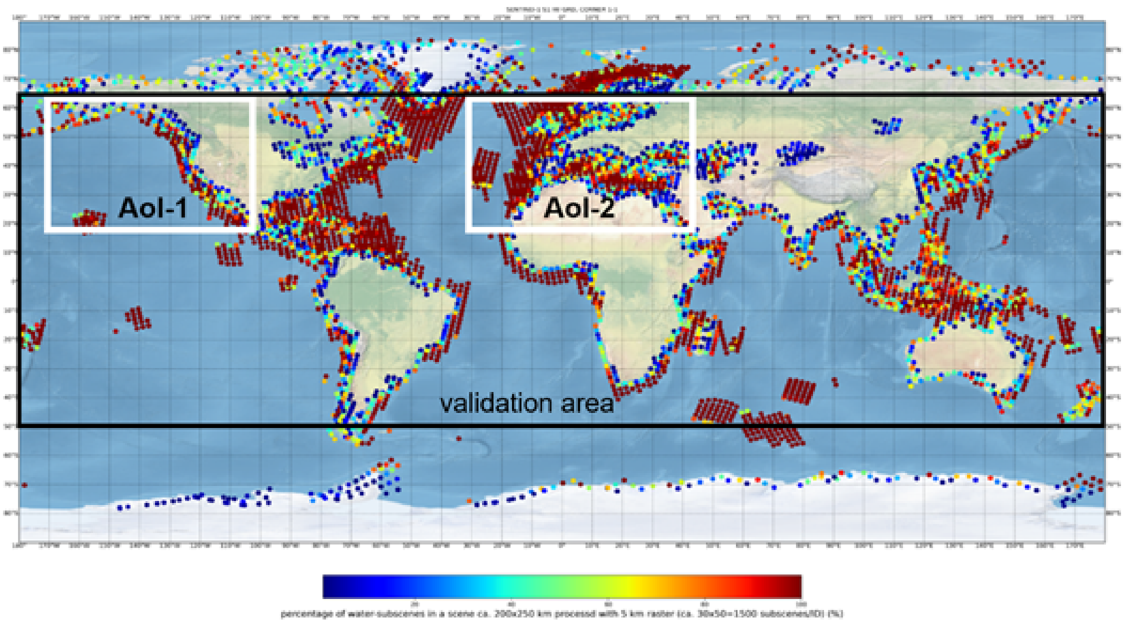

Figure 12 : Example coverage of marine scenes S1 IW in 2020. Shown are upper-left corner coordinates of each ID product, color shows percentage of processed points (to filtered out and land) for each ID.

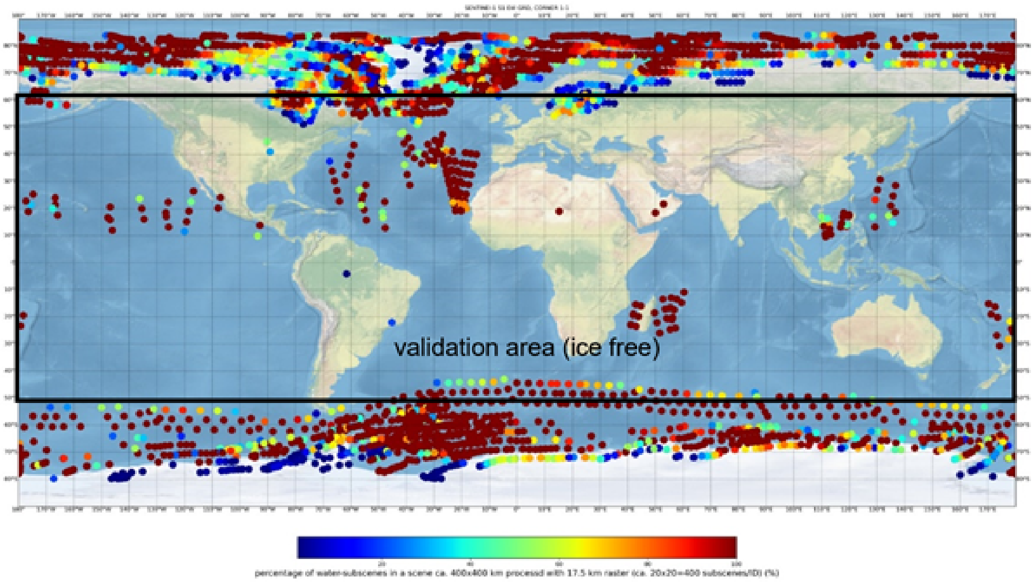

Figure 13 : Example coverage of marine scenes S1 EW in 2020. Shown are upper-left corner coordinates of each product ID, color shows percentage of water points (to filtered out and land) for each ID.

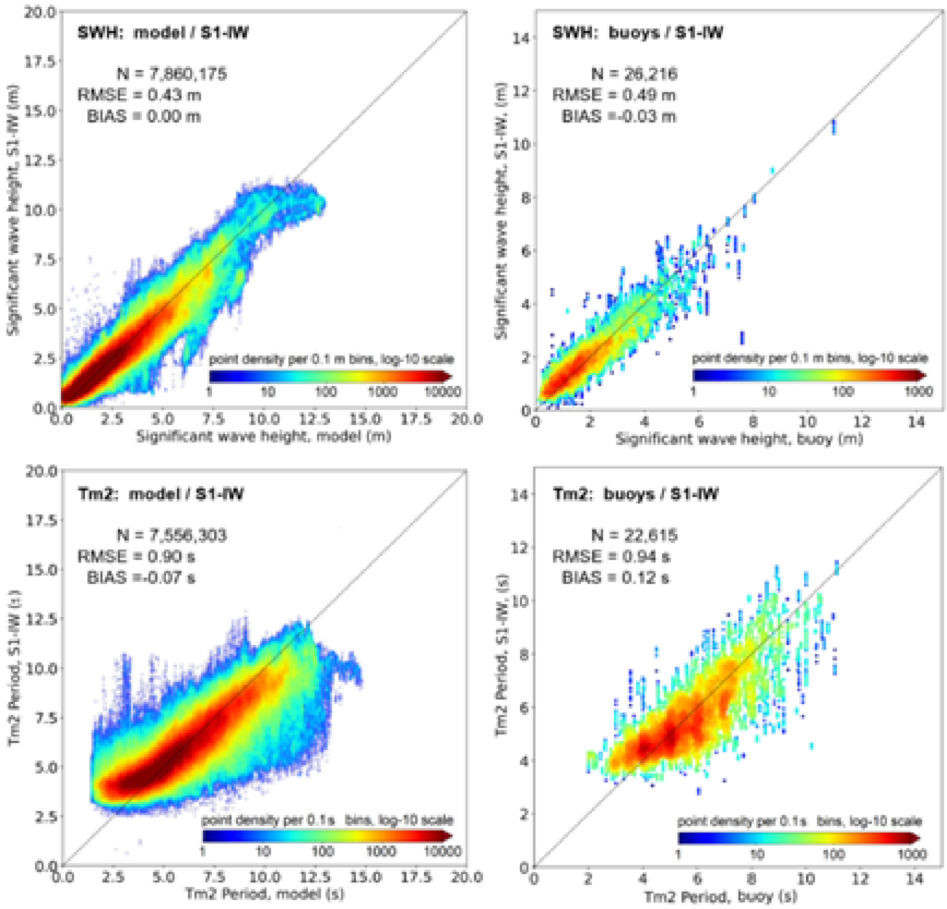

Figure 14 : S1 IW comparisons for total significant wave height SWH and second moment wave period Tm2 estimated using SAR-SeaStaR with model and buoys (10 km collocations) 2 AoIs.

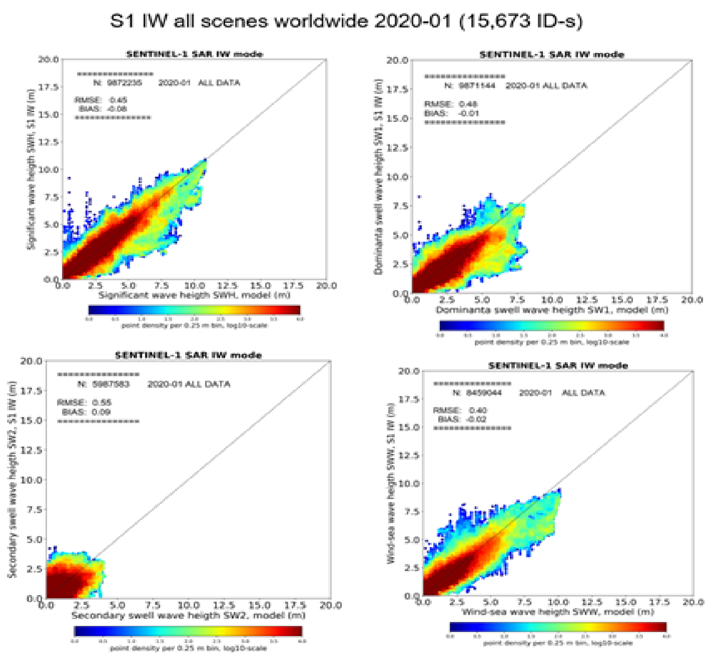

Figure 15 : Validation all worldwide acquired S1 IW with MFWAM (CMEMS) for January 2020 for total SWH, swell and windsea components (non-zero values, -50°<LAT<60°). There are ~16,000 ocean scenes with ~6M samples, processing raster of 5 km.

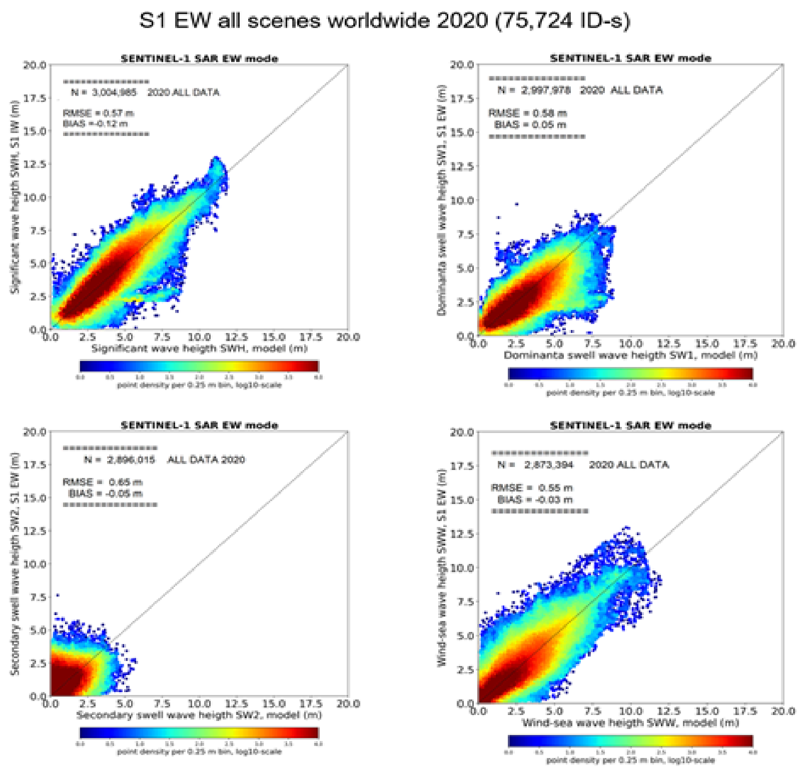

Figure 16 : Validation all worldwide acquired S1 EW scenes with MFWAM (CMEMS) in 2020 for total SWH, swell and windsea components (non-zero values, -50°<LAT<60°). There are 75,000 ocean scenes with ~3M samples, processing raster of 17.5 km.

Table 2 : DLR S1 IW SWH RMSE distribution for three source combinations (DLR-S1, model, buoys) for four sea state domains

SWH RMSE (m) distribution DLR all data 2020 2 AoIs*							
Wave height domain (m)	model – S1 IW (m)	buoys – S1 IW (m)	buoys – model (m)	N model	% model	N buoys	% buoys
0.0 < SWH< 1.5	0.330	0.405	0.209	20,998,132	48.85	55,029	46.89
1.5 < SWH< 3.0	0.393	0.464	0.315	14,675,728	34.14	43,098	36.72
3.0 < SWH< 6.0	0.543	0.653	0.502	6,444,016	14.98	17,734	15.12
6.0 < SWH	0.835	1.358	0.973	873,876	2.03	1,490	1.27
TOTAL	**0.41**	**0.484**	**0.327**	**42,991,725**	**100%**	**117,351**	**100%**

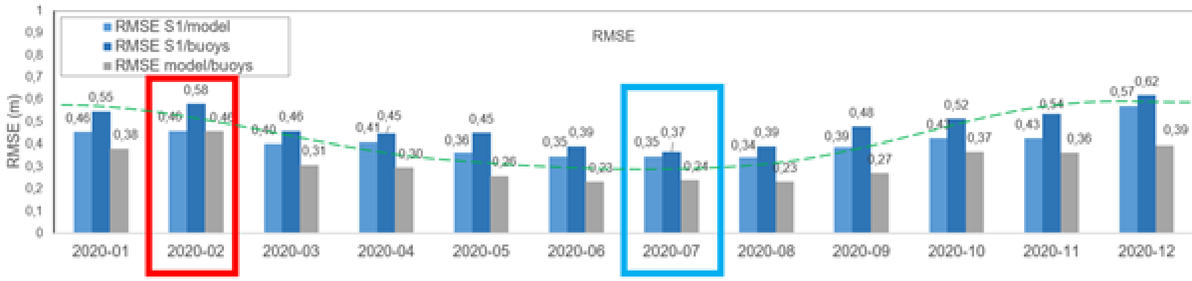

Figure 17 : Monthly RMSEs for three pairs of data (DLR_S1_SWH, model, buoys) during the year 2020. The tendency (green dotted line) can be seen with lower RMSE in summer calm months.

2.3.3 Cross comparison and uncertainties between SAR, model and buoys

The cross-comparison for three combinations of SWH data sources: DLR_OCN estimated from SAR, buoys and MFWAM model (only buoy locations) were carried out. The RMSEs were estimated in total and for four SWH domains. Figure 18 visualizes the results. The uncertainty for all sources increases with the wave height. The averaged extrapolated uncertainty can be assessed as *uncertainty*=0.20+0.10·SWH (i.e. 20 cm ground noise with additional 10% of SWH value).

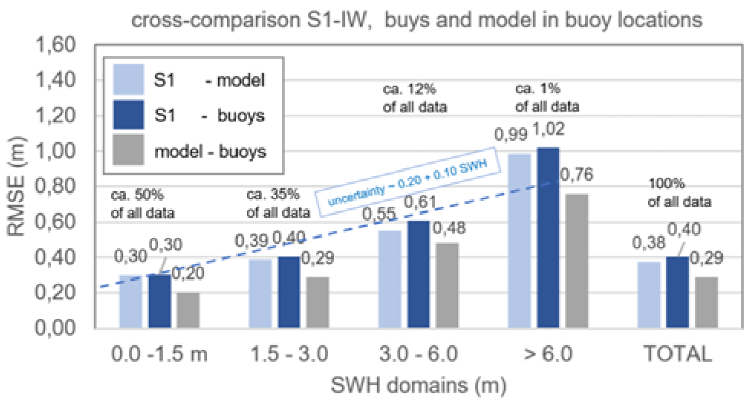

Figure 18. Cross comparison of SWH for DLR_OCN, buoys and model (only buoy locations). The uncertainty for all sources grows with higher value of the wave height. The averaged extrapolated uncertainty can be assessed as uncertainty=0.20+0.10·SWH.

2.3.4 SAR-derived wave partition parameters

For Phase 2 of the Sea State CCI+ project, ODL is currently developing a deep learning based approach [Goodfellow et al 2016., Reichstein et al. 2019,Camps-Valls et al. 2021, Quach et al. 2020] to produce and incorporate wave partition parameter estimations from SAR data into the CCI+ Sea State data product.

The deep learning based algorithm under development focuses directly on the inversion of directional wave spectra from SAR cross-spectra, with the computation of wave partition parameters from the obtained directional wave spectra to be achieved by direct partitioning of the produced directional wave spectra, using a watershed partitioning algorithm, followed by the computation of partition parameters for each detected partition.

As such, the proposed algorithm aims at leveraging machine learning and deep learning techniques to improve on the current model-based inversion, and perform a fully data-driven inversion of the wave directional spectra from SAR cross-spectra. The final objective is thus to provide **estimations** of **wave partition parameters**, including **partition significant wave height** ($H_s^{partition\text{-}n}$), **partition peak wavenumber** ($K_{peak}^{partition\text{-}n}$) and **azimuth**($\theta_{peak}^{partition\text{-}n}$), and **spectral** ($K_{sp}^{partition\text{-}n}$) and **directional spread** ($\theta_{sp}^{partition\text{-}n}$) from SAR imagery, in particular for the wave modes of Sentinel-1, ERS and Envisat.

From a methodological point of view, recent developments focused on understanding and modelling uncertainty sources in deep learning approaches, as well as incorporating uncertainty estimation directly into deep learning frameworks (such as Bayesian neural networks [Goan et al. 2020], ensemble methods [Dietterich 2000], test-time data augmentation [Shorten et al. 2019], amongst others) have shown promising results, which makes them particularly appealing as a relevant tool to help quantify and potentially improve the CCI+ Sea State data product uncertainty budget.

Nonetheless, the inclusion of uncertainty estimation within neural network based approaches requires either:

- modifying the neural network architecture and training procedure, which can potentially impact network performance and generalization properties
- exploiting multiple concurrent models, trained with different initializations and parametrizations, to produce an ensemble of outputs on which statistics can be computed, which requires considerable computational and post-processing time
- leverage hybrid approaches that rely on carefully chosen a priori or post-hoc statistical assumptions on network parameters to derive uncertainty estimates

Besides these issues, it must also be noted that additional uncertainty sources in the processing chain, including partitioning and partition cross-assignment associated errors, will also impact the final uncertainty budget of the proposed methodology.

Currently, the deep learning based approaches proposed for wave spectra inversion from SAR cross-spectra remain at an exploratory phase, and still require additional efforts to achieve the maturity required to adequately implement these modifications.

Exploring these approaches is of paramount importance for the characterization of uncertainty for SAR-derived partition parameters, and further efforts are to be performed once a sufficient maturity level for the proposed methodology is achieved. In particular, given the differences in mission and sensor characteristics between the considered SAR datasets and missions (resolution, noise levels, image size, etc), these approaches will help quantify the impact of such differences in wave partition parameter retrieval. Indeed, ensuring consistency between SAR imaging missions for wave partition parameter retrieval will require the re-adaptation of the proposed methodology to better suit the individual characteristics of each SAR imaging mission, which highlights the importance of a thorough characterization and comparison of the uncertainty budget for the different SAR missions to be processed and included in the CCI+ Sea State data archive.

As previously explained, given that the algorithm is still under development, a thorough evaluation of uncertainty and error propagation has not been completed at this stage, but will be performed once the final version of the proposed methodology is implemented, for both internal verification and validation purposes. Preliminary internal verification results, however, provide a current estimate of the algorithm's performance. As such, internal verification statistics can be regarded as an adequate proxy for uncertainty at this early stage, provided that they are computed on ground truth data from the same source used as groundtruth during training (in this case, numerical model outputs from WaveWatch III). It should be noted that these results reflect the performance of the current best version of the algorithm, and are expected to be improved with future versions.

During the development stage, the base algorithm for directional wave spectra inversion from SAR images was trained on an experimental dataset of colocalized Sentinel-1 A/B VW1 & VW2 cross-spectra and WaveWatch3 wave directional spectra, produced at Ifremer in the context of the SARWAVE project.

The complete SARWAVE dataset comprises 2.202.370 pairs of colocalized SAR Sentinel-1 A/B VW1 & VW2 cross-spectra and WaveWatch3 wave directional spectra, spanning the whole global ocean from January 2015 to October 2013.

A subset of 180.000 randomly sampled pairs in -55°<LAT<60° (to avoid any potential issues related to ice coverage) was used to build the experimental dataset used during development. The Sentinel 1 SAR cross-spectra are subsequently interpolated onto the WaveWatch3 spatiotemporal grid, with the following characteristics:

- Linear sampling in direction: $N_\theta = 72$ equidistantly spaced directions (5° bins)
- Logarithmic sampling in wavenumber: $N_k = 60$ wavenumber values sampled between minimum wavelength $L_{min} = 30$ (corresponding to $k_{max} = 0.2094$) and maximum wavelength $L_{max} = 1200$ (corresponding to $k_{min} = 0.005235$), to achieve optimal sampling of the energetic part of the SAR spectra:
  - $k_i = \alpha^i k_{min}$ with $\alpha = \left(\frac{k_{max}}{k_{min}}\right)^{\frac{1}{N_{k-1}}}$, $k_{min} = \frac{2\pi}{L_{min}}$ and $k_{max} = \frac{2\pi}{L_{max}}$

The experimental dataset was subsequently split randomly into a training dataset (64% of the experimental dataset) for model training, a validation dataset (16% of the experimental dataset) for in-training validation and performance monitoring, and a test dataset (20% of the experimental dataset).

For illustration purposes, we present preliminary results obtained from the current version of the proposed machine learning model, trained using the train and validation datasets and applied to the test dataset. These preliminary results provide a first estimate of the current algorithm's performance but, as previously stated, are expected to be improved upon finalizing the development of the proposed methodology. A through analysis of the uncertainty and error propagation for the proposed methodology, including a comparison to wave partition parameters derived from both numerical model and in-situ observations, as well as a comparison with the other SAR processing algorithms for global significant wave height SWH and mean wave period T_{m01}, will also be performed once the development stage is completed.

2.3.5 Preliminary results

All preliminary results presented below were computed on the full test dataset, which excludes, by definition, observations at -55°<LAT<60° to avoid ice coverage. For uncertainty characterization, both global and partition parameters are computed for the wave directional spectra obtained using the proposed neural network based inversion algorithm, as well as for the corresponding colocalized WaveWatch3 directional spectra, which provides the ground truth with respect to which relevant uncertainty metrics are computed. For each considered parameter, scatterplots, error histograms and geographical error maps are presented.

Given that the predicted directional wave spectra and the corresponding colocalized WaveWatch3 wave spectra are partitioned independently (via a watershed partitioning scheme), comparison of partition parameters requires the matching and cross-assignment of identified partitions.

In this regard, the partition matching scheme used to produce the results presented below relies on matching partitions with respect to the distance between their spectral peaks. In this way, for each pair of predicted and ground truth wave spectra, partitions whose peaks are closer together are considered as a match. This procedure is performed in ascending order with regards to peak distance, so that pairwise peak distances between all identified partitions are first computed, and partition pairs with the closest peaks are matched first. To further avoid possible errors due to the independent partitioning, only matches involving partitions that intercept are retained. Once matched, partitions are flagged to avoid double matches, and the procedure is performed iteratively until all partition pairs have been processed. It should be noted that the chosen cross-assignment methodology may sometimes produce orphaned partitions that are not matched, and are thus discarded.

Below, preliminary results for the following parameters are presented:

- Global parameters:
    - SWH

- Partition parameters:
    - SWH
    - Partition peak wavelength
    - Partition peak direction

For each parameter, we present scatterplots and error (predicted value - ground truth value) histograms comparing parameters computed from directional wave spectra obtained from SAR cross spectra with the proposed machine learning methodology vs parameters computed from the corresponding colocalized WaveWatch3 directional wave spectra.

Parameter RMSE and bias geographic distribution plots are also depicted. Importantly, colorbars for geographic distribution plots include, at their tops and bottoms, both the chosen colormap limits (closer to the colorbar), as well as the observed maximum and minimum values (further from the colorbar).

Moreover, similarly to results presented previously, the same analysis is performed for four different sea states domains computed with respect to both global and partition SWH values.

For interpretation and analysis purposes, observation count geographical distribution plots are also included for global parameters, and for sea states domains computed with respect to global and partition SWH values, at the beginning of each corresponding section.

Finally, a table summarizing all presented results is included at the end of each section.

2.3.5.1 Global parameters: Significant Wave Height (SWH)

Global analysis:

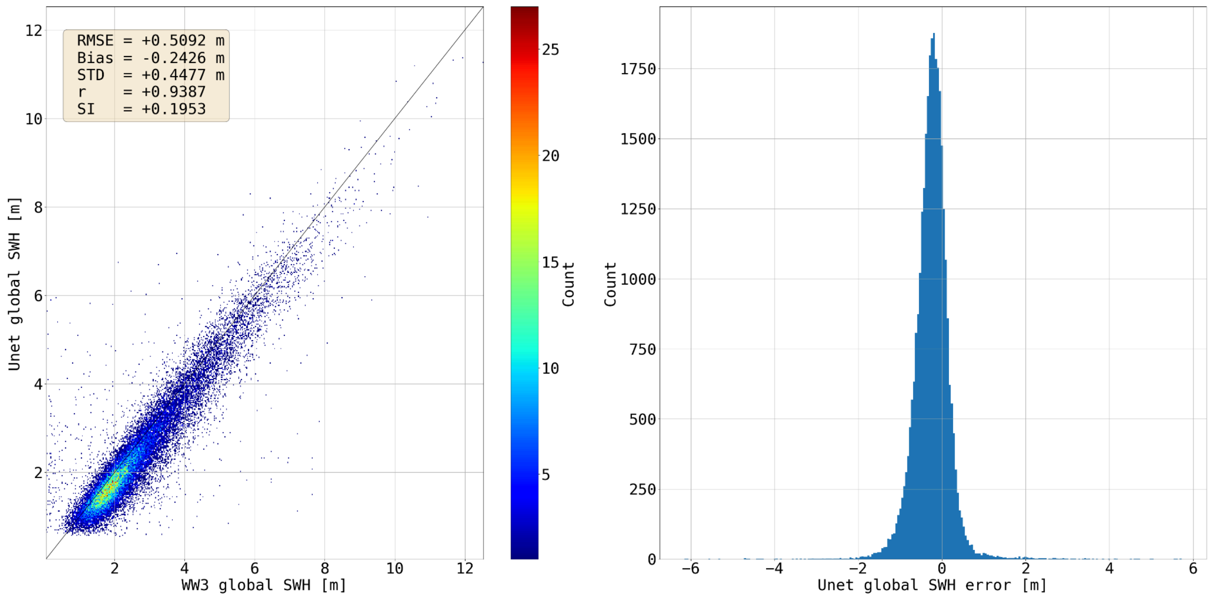

Figure 18: [Left] Comparison of global SWH: U-net based directional wave spectra global SWH vs Groundtruth colocalized WaveWatch3 directional wave spectra global SWH. [Right]

Global SWH error histogram: U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

From Figure 18, a mean global SWH RMSE value of around 0.5 m is observed. Clearly, global SWH retrieval performance needs to be further improved, as it is not currently at levels obtained with alternative algorithms, such as Stopa or DLR. Importantly, we believe these high RMSE values to be closely related to the observed global bias of around -0.25 m. This is an issue we are currently working on, and believe that correcting this negative bias should allow us to obtain acceptable global SWH retrieval performance and a mean global SWH RMSE value comparable to those obtained with alternative algorithms

As discussed during Progress Meeting 3, we believe these issues to be related to the current methodology's underestimation of partition energy levels, which we hypothesise to be related to both the normalization/standardization scheme used on training data, as well as to possible issues with the parametrization of the neural network's last layer. Tests are currently underway to improve on these issues and further improve global SWH retrieval accuracy.

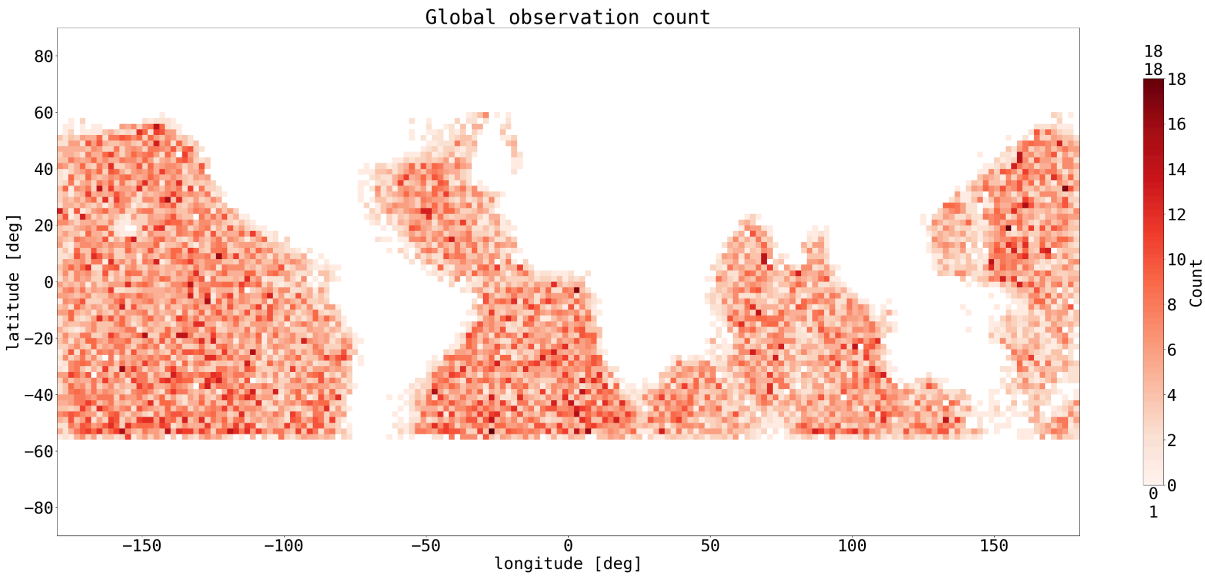

Figure 19: Global observation count geographical distribution.

Figure 19 clearly demonstrates an homogenous geographical sampling that should not introduce any considerable geographical biases.

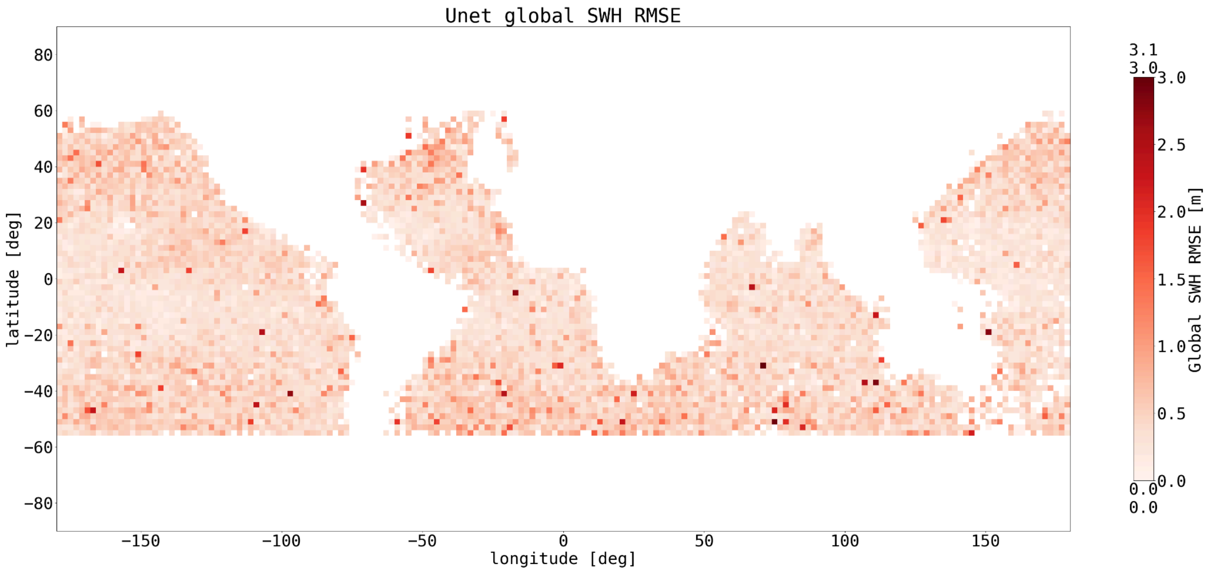

Figure 20: Global SWH RMSE geographical distribution: U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

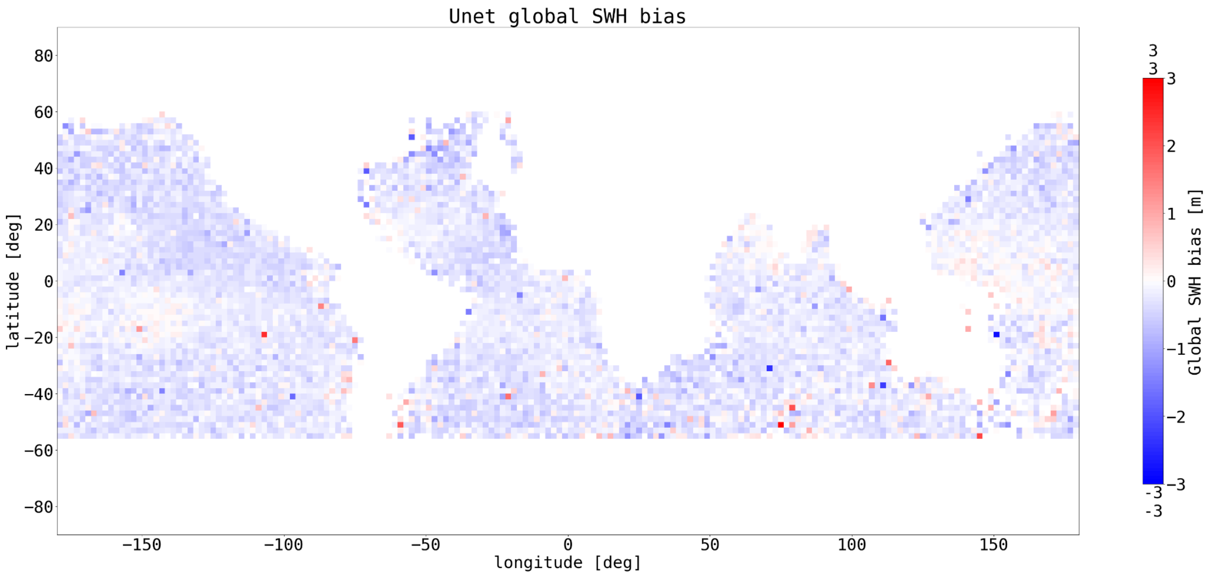

Figure 21: Global SWH bias geographical distribution: U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

Figures 20 and 21 seem to indicate that neither RMSE or bias present any geographical biases or patterns, with a relative homogeneous geographical distribution of values and outliers, which further supports our hypothesis of global SWH retrieval performance issues being related to data normalization/standardization and/or to possible parameterization issues within the neural network's last layer.

Sea states domains analysis (with respect to global SWH):

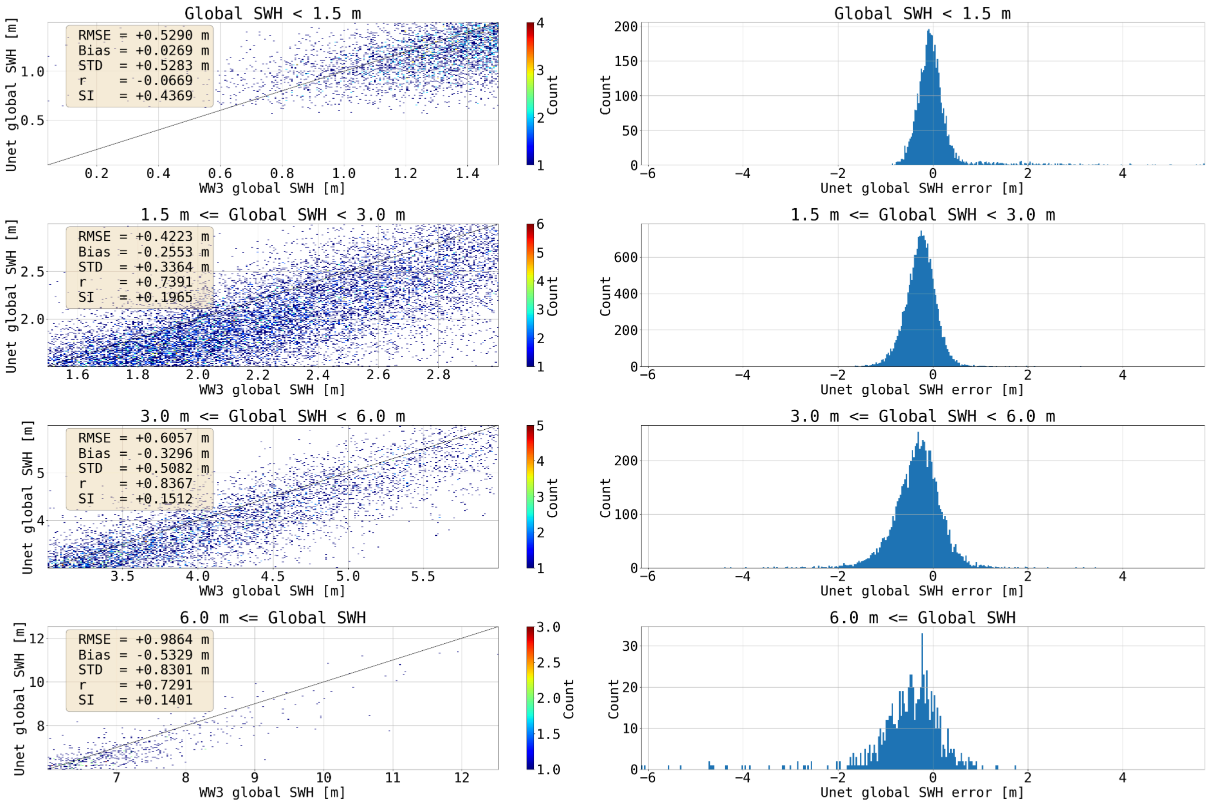

Figure 22: [Left] Comparison of global SWH for four different sea state domains (w.r.t. global SWH): U-net based directional wave spectra global SWH vs Groundtruth colocalized WaveWatch3 directional wave spectra global SWH. [Right] Global SWH error histograms for four different sea state domains (w.r.t. global SWH): U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

From Figure 22, global SWH retrieval performance peaks for moderate SWH levels (between 1.5m and 3.0m), and degrades as we approach more extreme sea states domains, particularly for high SWH values. In particular, this is partially explained by the fact that there is a considerably higher number of data points within the moderate SWH levels, which means the data-driven model will be better adapted to handle such cases. This could possibly be improved by re-equilibrating the training dataset in pre-processing, applying an histogram normalization scheme to the training dataset, weighting observations by sea state domain during training, or further modifying the loss function to take sea state domain imbalance into account.

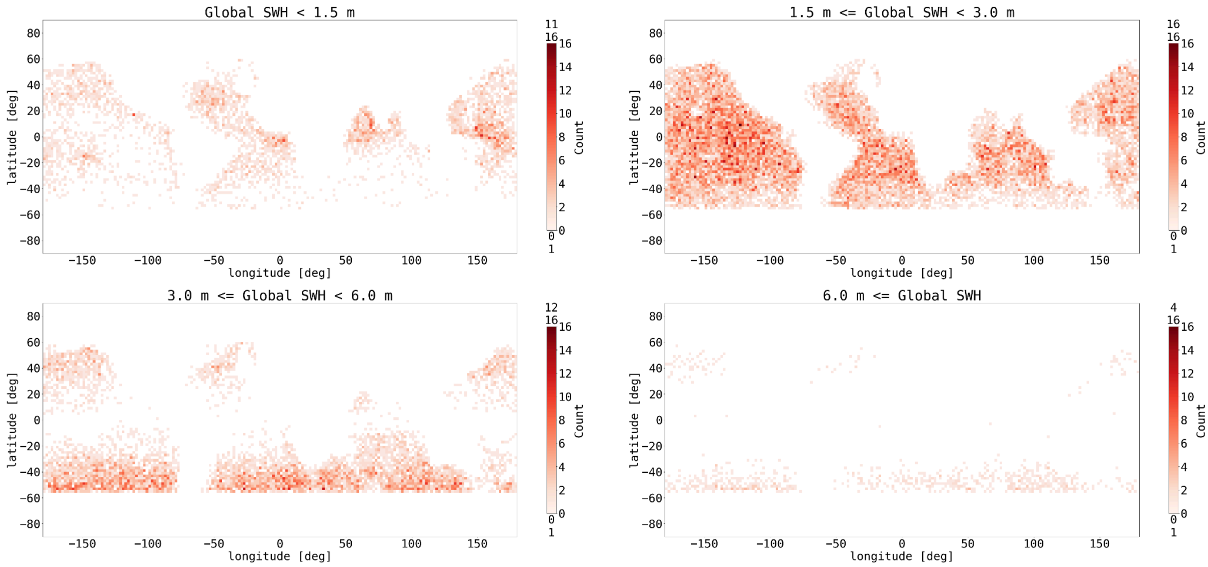

Figure 23: Observation count geographical distribution for four different sea state domains (w.r.t. global SWH).

Figure 23 shows clear geographical patterns for different sea state domains (w.r.t. global SWH), with a rather homogeneous geographical distribution of moderate sea states (global SWH between 1.5m and 3.0m), low global SWH values being more present near the Equator and more energetic sea states being more predominant in the southern ocean.

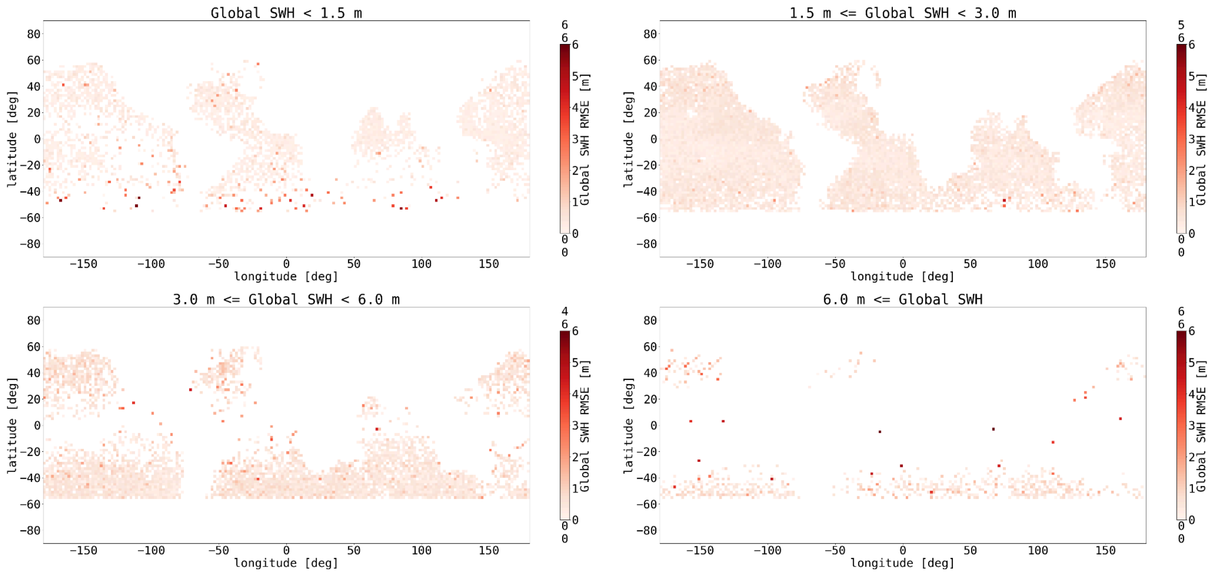

Figure 24: Global SWH RMSE geographical distribution for four different sea state domains (w.r.t. global SWH): U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

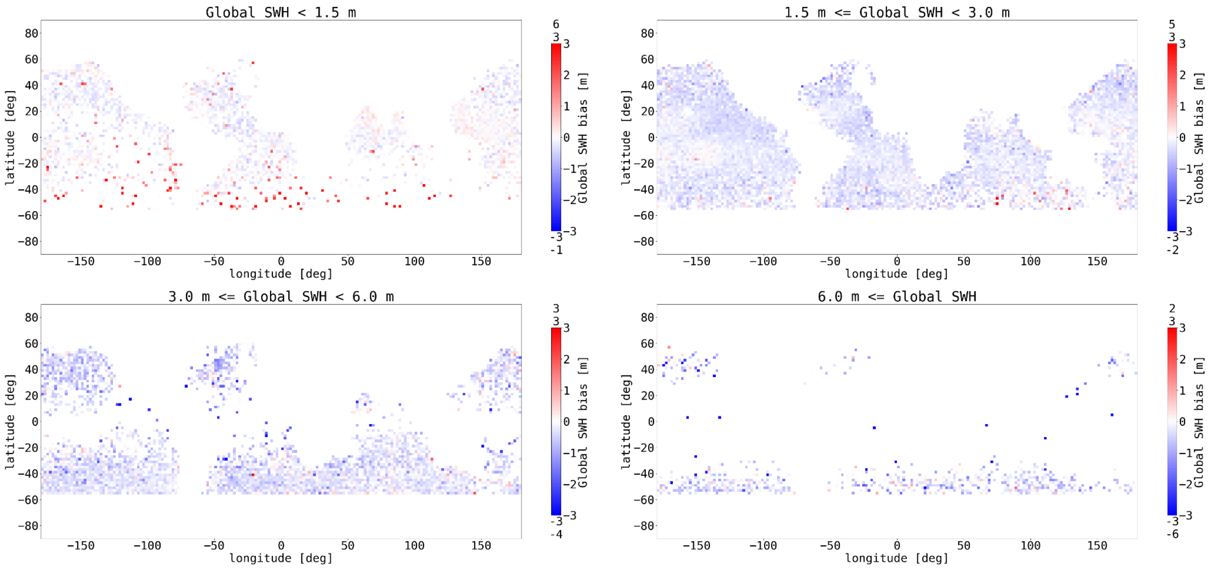

Figure 25: Global SWH bias geographical distribution for four different sea state domains (w.r.t. global SWH): U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

Despite the clear observation distribution geographical patterns depicted in Figure 23, Figures 24 and 25 show a rather homogeneous geographical distribution of global SWH RMSE and bias values, with the exception of low energy sea states (global SWH<1.5m), which appears to present a considerably lower global SWH retrieval performance towards the southern seas, which may be partially related to the low density of observations for these sea states in these regions. Moreover, RMSE and bias distribution for moderate sea states (global SWH between 1.5m and 3.0m) seems coherent with the geographical distribution of observations for such sea states. Finally, outliers seem to be homogeneously distributed for all sea states domains, and seem to be more prevalent towards both low observation density zones as well as coastal areas. Once again, these findings further support our hypothesis of global SWH retrieval performance issues being related to data normalization/standardization and/or to possible parameterization issues within the neural network's last layer.

Results summary:

Table 3 : ODL Global SWH preliminary results summary

Global SWH							
Wave height domain (m)							
	RMSE	Bias	STD	r	SI	Count	Data percentage
Global	0.51 m	-0.24 m	0.45 m	0.94	20%	31074	100%
0.0 < global SWH< 1.5	0.53 m	+0.03 m	0.53 m	0.67	44%	4209	13.55%
1.5 < global SWH< 3.0	0.42 m	-0.26 m	0.34 m	0.74	20%	18364	59.10%
3.0 < global SWH< 6.0	0.61 m	-0.34 m	0.51 m	0.84	15%	7701	24.78%
6.0 < global SWH	0.99 m	-0.33 m	0.83 m	0.73	14%	800	2.57%

2.3.5.2 Partition parameters : *Significant Wave Height (SWH):*

Global analysis:

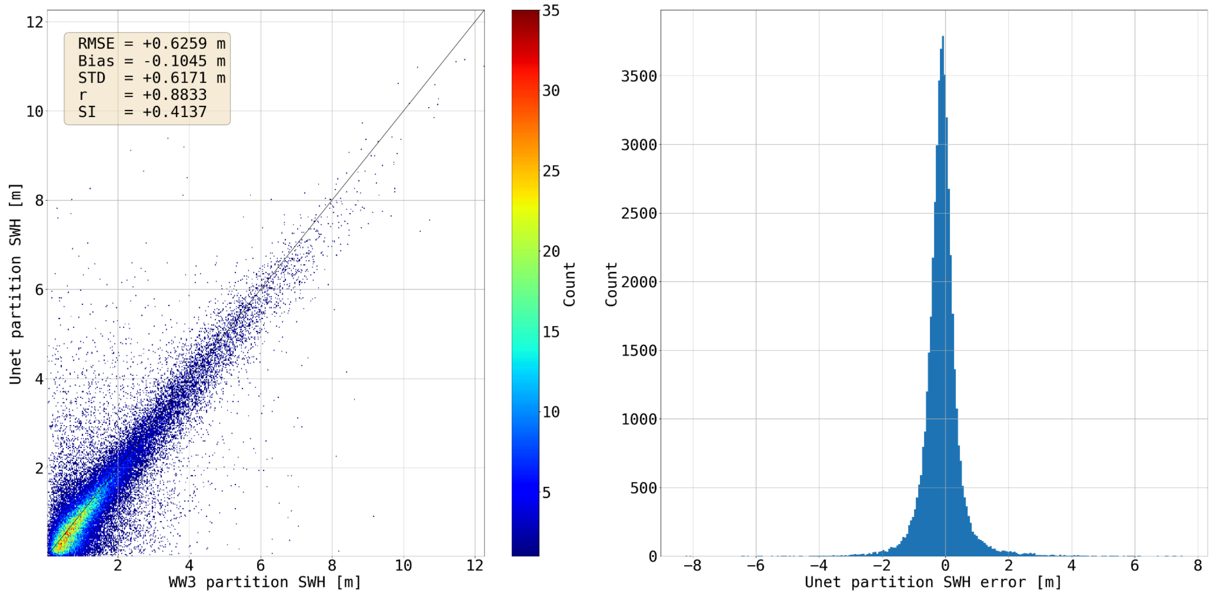

Figure 26: [Left] Comparison of partition SWH: U-net based directional wave spectra partition SWH vs Groundtruth colocalized WaveWatch3 directional wave spectra partition SWH. [Right] Partition SWH error histogram: U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

Figure 26 indicates a global mean RMSE value of 0.63 m and a global mean bias of -0.11 m for the retrieval of partition SWH. As with global SWH, current efforts are underway to further improve partition SWH retrieval performance, which we also believe to be associated to the current methodology's underestimation of energy at the partition level, and thus related to data normalization/standardization and/or to possible parameterization issues within the neural network's last layer.

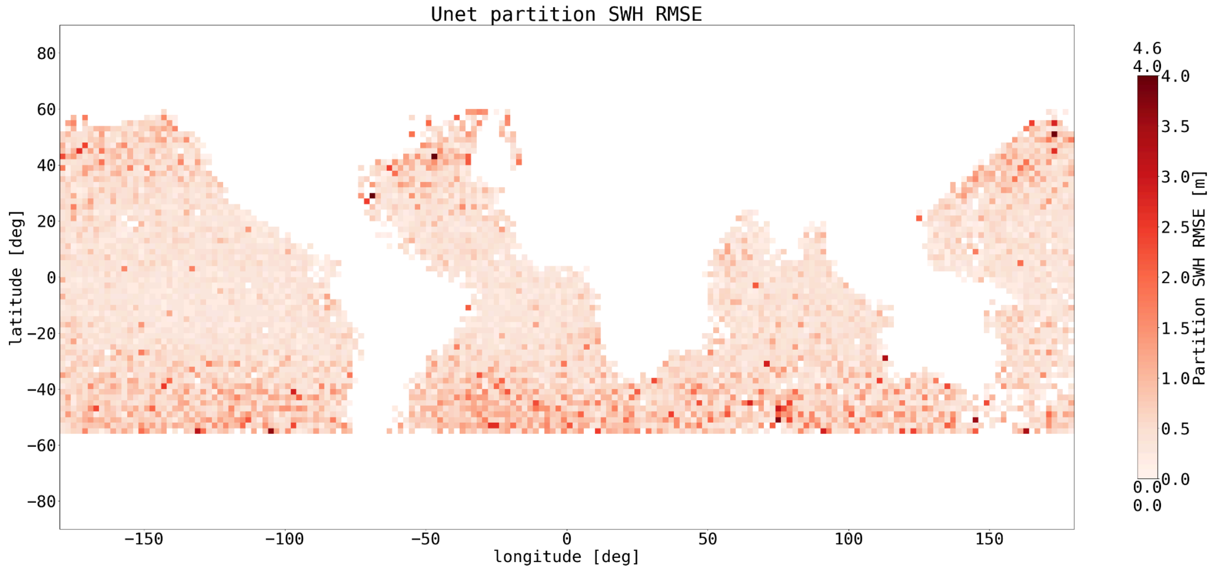

Figure 27: Partition SWH RMSE geographical distribution: U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

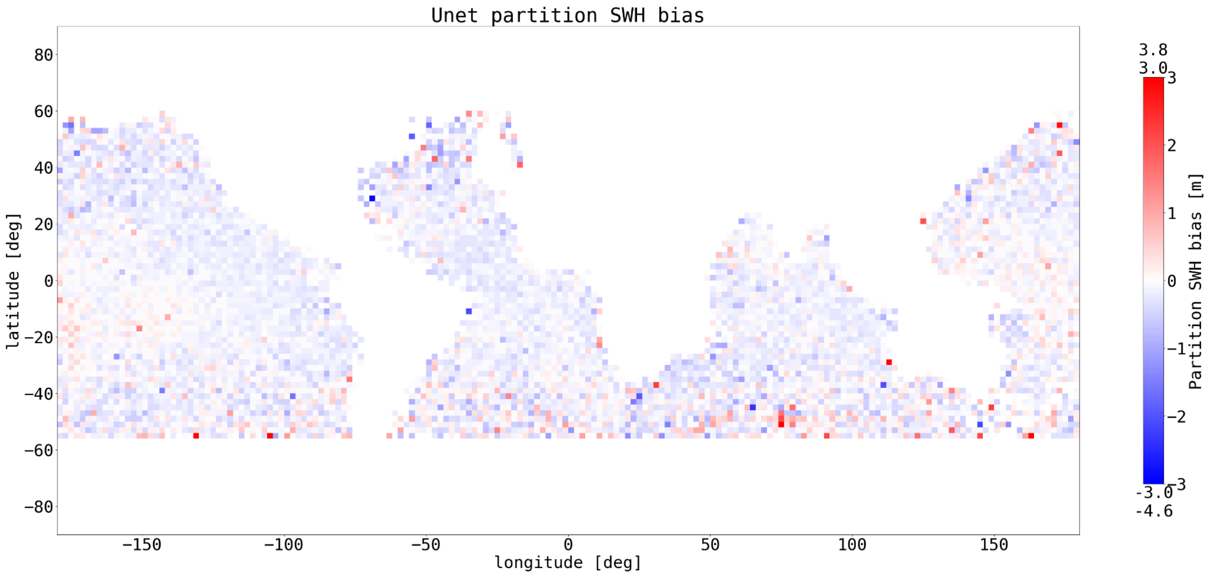

Figure 28: Partition SWH bias geographical distribution: U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

Regarding the geographical distribution of partition SWH RMSE and bias, Figures 27 and 28 depict results similar to those obtained for global SWH retrieval, with a rather homogeneous distribution of lower RMSE and bias values across the global ocean. However, differently from global SWH, a concentration of higher RMSE and bias values towards the southern oceans, in particular, but also towards the northern ocean and some coastal areas, is observed. In particular, these results may be partially explained by the predominance of lower energy sea states (w.r.t. partition SWH) within the dataset (as indicated from Table 4), which present clear geographical distribution patterns (Figure 30) matching the global geographical distribution of partition SWH RMSE and bias.

Sea states domains analysis (with respect to partition SWH):

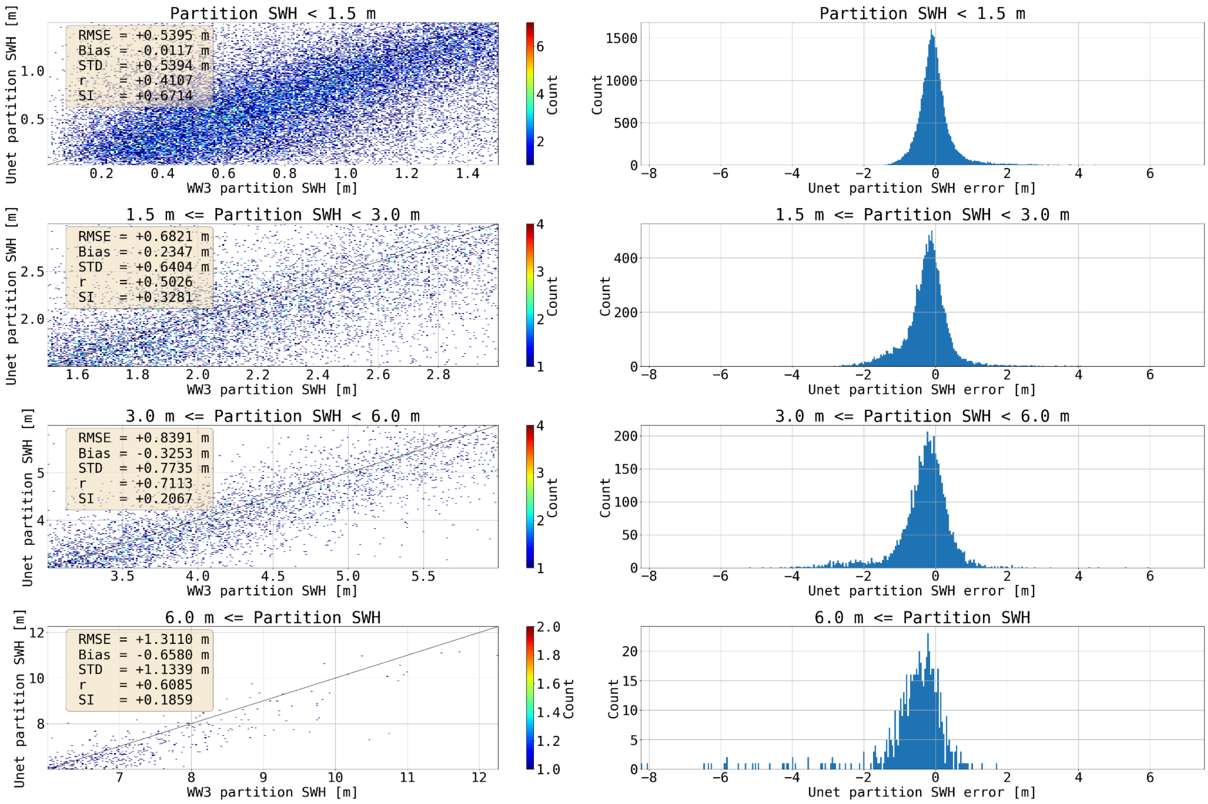

Figure 29: [Left] Comparison of partition SWH for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra partition SWH vs Groundtruth colocalized WaveWatch3 directional wave spectra partition SWH. [Right] Partition SWH error histograms for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

Regarding the distribution of partition SWH retrieval performance at different sea state domains (w.r.t. partition SWH), from Figure 29 it seems clear that the majority of the retrieved partitions lie within the less energetic sea state domains (with 65% of the retrieved partitions in the lower SWH tier (partition SWH < 1.5 m) and 23% of the retrieved partitions in the moderate SWH tier (1.5 m < partition SWH < 3.0 m), as indicated in Table 4). As expected, lower RMSE and bias values are associated with sea states with higher observation density, so that partition SWH retrieval performance is best for the less energetic sea state domain (partition SWH < 1.5 m), and deteriorates for more energetic sea state domains.

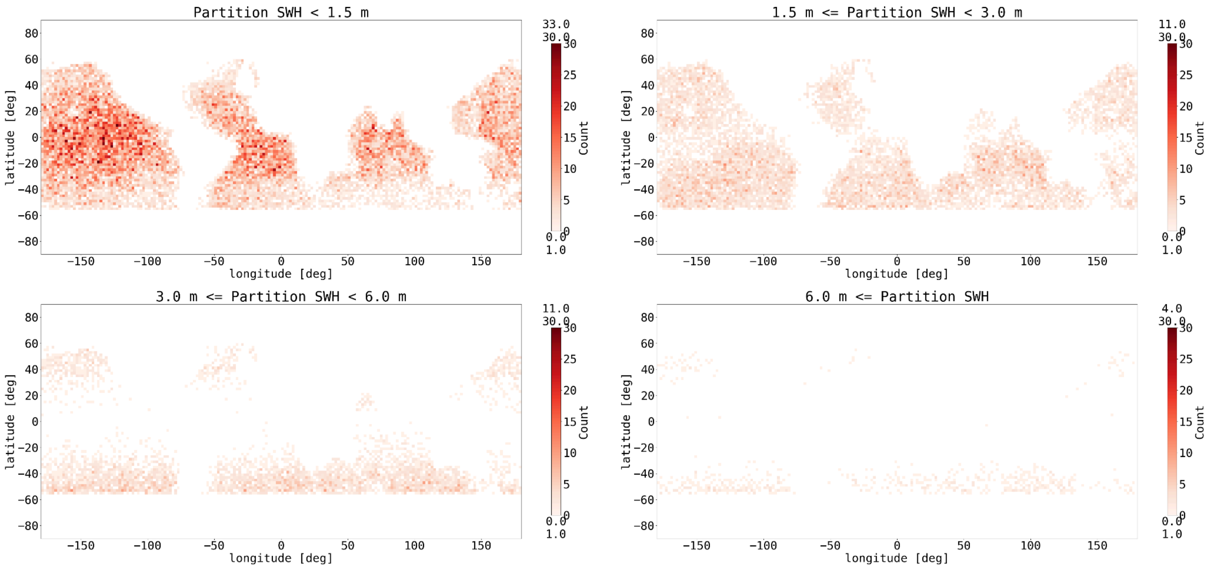

Figure 30: Observation count geographical distribution for four different sea state domains (w.r.t. partition SWH).

The geographical distribution of the four considered sea state domains (w.r.t. partition SWH) is depicted in Figure 30, with a clear predominance of low energy sea states (partition SWH < 1.5 m) for -40°<LAT<40°, an homogeneous distribution of moderate energy sea states (1.5 m < partition SWH < 3.0 m), and higher energy sea states more predominant toward the southern ocean.

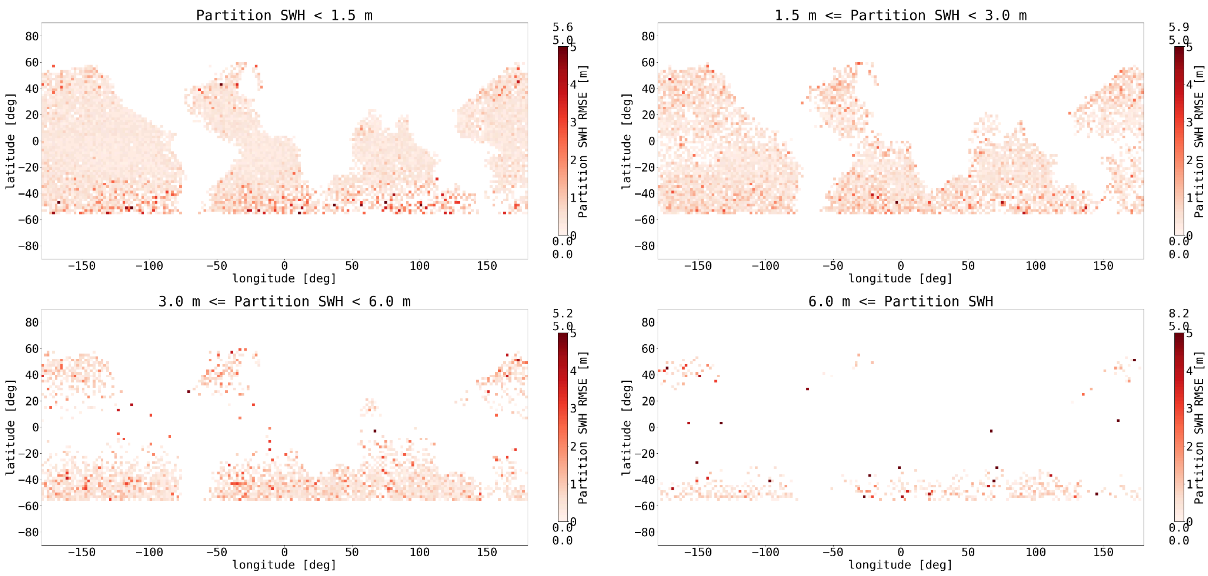

Figure 31: Partition SWH RMSE geographical distribution for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

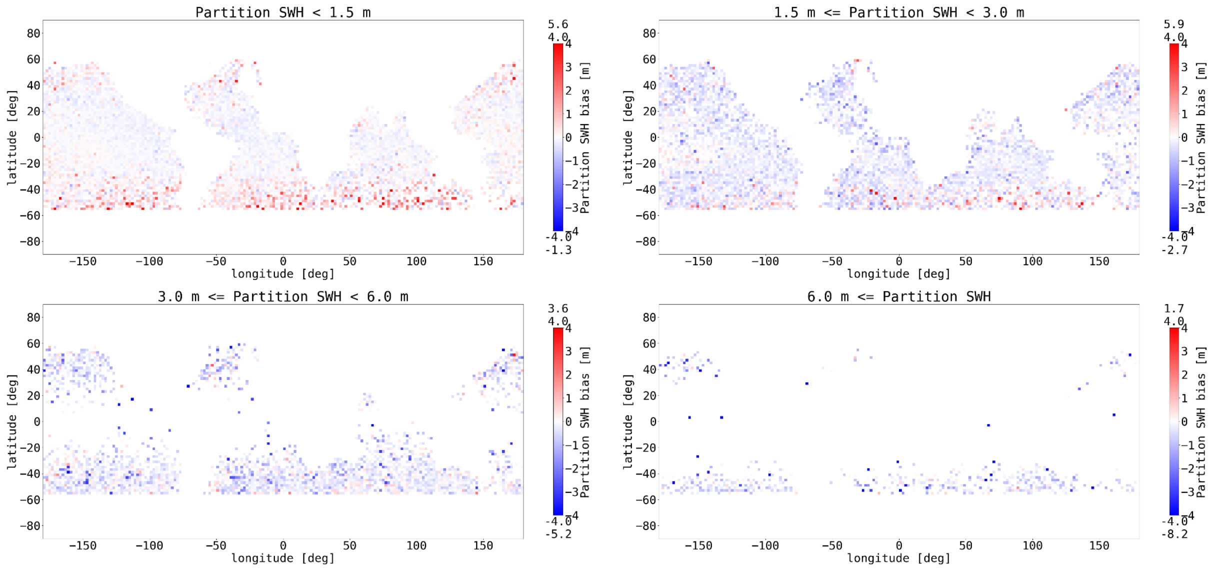

Figure 32: Partition SWH bias geographical distribution for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra vs Groundtruth colocalized WaveWatch3 directional wave spectra.

As far as the geographical distribution of partition SWH RMSE and bias for different sea state domains (w.r.t. partition SWH) is concerned, it may be noted in Figure 32 that the geographical distribution of RMSE and bias for the less energetic sea state domain (partition SWH < 1.5 m) closely matches the global geographical distribution of partition SWH error and bias (Figures 27 and 28), with higher RMSE and bias values more predominant towards the south.This is to be expected given the predominance of lower energy sea states belonging to this domain within the dataset (with 65% of observations belonging to this sea state domain). A similar, yet weaker, geographical distribution pattern can be observed for moderate energy sea states (1.5 m < partition SWH < 3.0 m), whereas RMSE and bias metrics seem to distribute homogeneously for higher energy sea states, with outliers more predominant in lower observation density areas and coastal zones.

Results Summary:

Table 4 : ODL Partition SWH preliminary results summary

Partition SWH							
Wave height domain (m)							
	RMSE	Bias	STD	r	SI	Count	Data percentage
Global	0.63 m	-0.11 m	0.62 m	0.88	41%	50038	100%
0.0 < partition SWH< 1.5	0.54 m	-0.01 m	0.54 m	0.41	67%	32421	64.79%
1.5 < partition SWH< 3.0	0.68 m	-0.23 m	0.64 m	0.50	33%	11949	22.88%
3.0 < partition SWH< 6.0	0.84 m	-0.33 m	0.77 m	0.71	21%	5062	10.12%
6.0 < partition SWH	1.31 m	-0.66 m	1.13 m	0.61	19%	606	1.21%

2.3.5.3 Partition parameters : Peak Wavelength (l_{peak})

Global analysis:

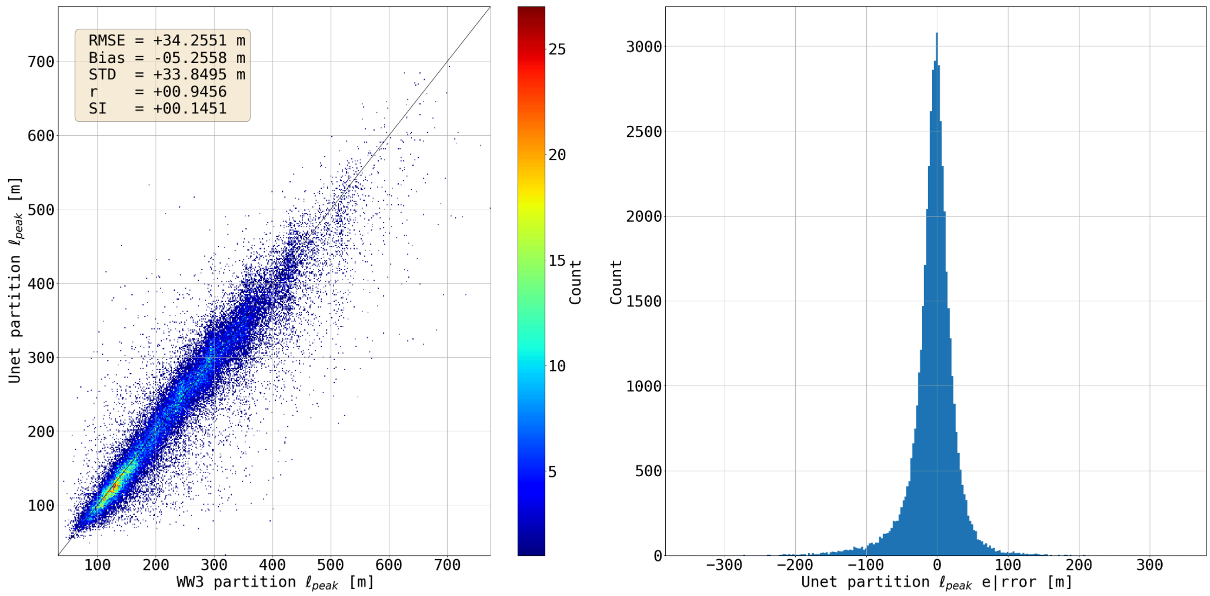

Figure 33: [Left] Comparison of partition peak wavelength: U-net based directional wave spectra partition peak wavelength vs Groundtruth colocalized WaveWatch3 directional wave spectra partition peak wavelength. [Right] Partition peak wavelength error histogram: U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

Regarding partition peak wavelength retrieval performance, a mean global RMSE of 34 m and a mean global bias of -5.3 m is observed in the scatterplot and error histogram depicted in Figure 33. As with partition SWH, current efforts aim at further improving the partition peak wavelength retrieval accuracy. This may be related, to some extent, to the level of output smoothing achieved by the proposed methodology, which directly relates to network architecture. As discussed during Progress Meeting 3, the proposed methodology may oversmooth outputs, which can be detrimental to peak wavelength retrieval accuracy. Improving on the compromise between output smoothing and robustness to observation noise and outliers can be achieved by further tuning network hyperparameters, and is expected to have a direct impact on partition peak wavelength retrieval performance.

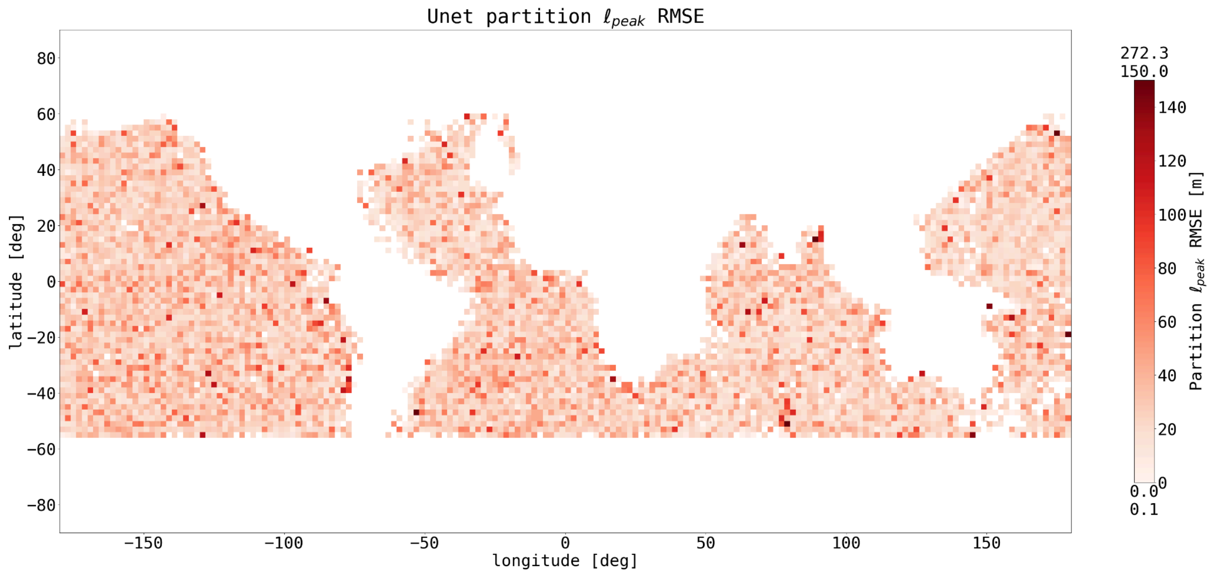

Figure 34: Partition peak wavelength RMSE geographical distribution: U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

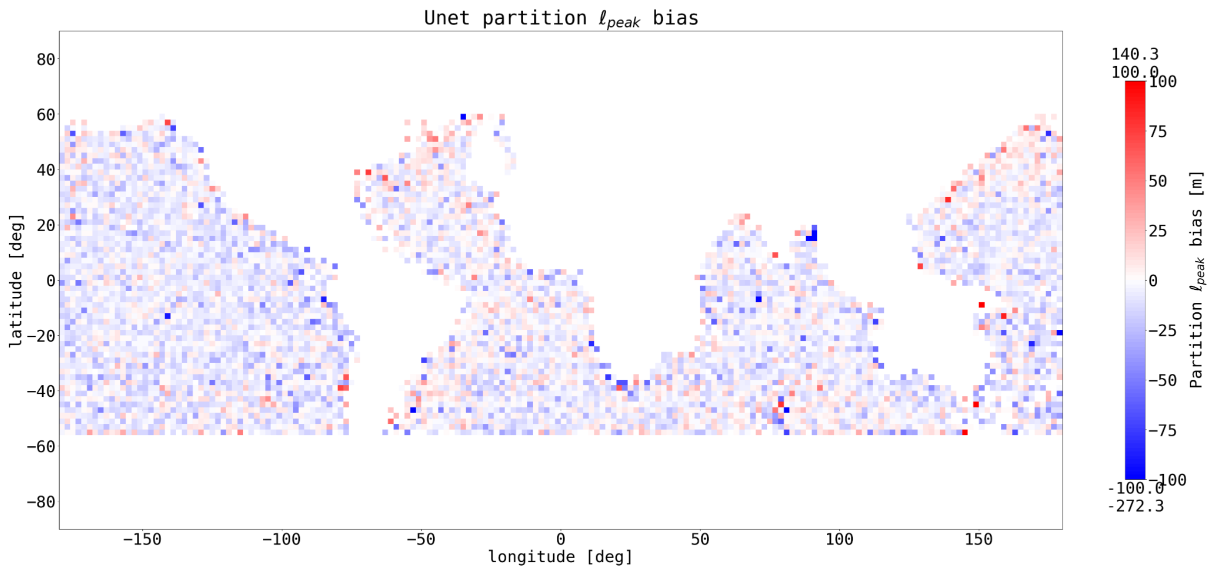

Figure 35: Partition peak wavelength bias geographical distribution: U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

As far as the geographical distribution of partition peak wavelength RMSE and bias metrics is concerned, depicted in Figures 34 and 35, no clear patterns of tendencies appear, with an homogeneous distribution of both metrics and associated outliers across the global ocean.

Sea state domains analysis (with respect to partition SWH):

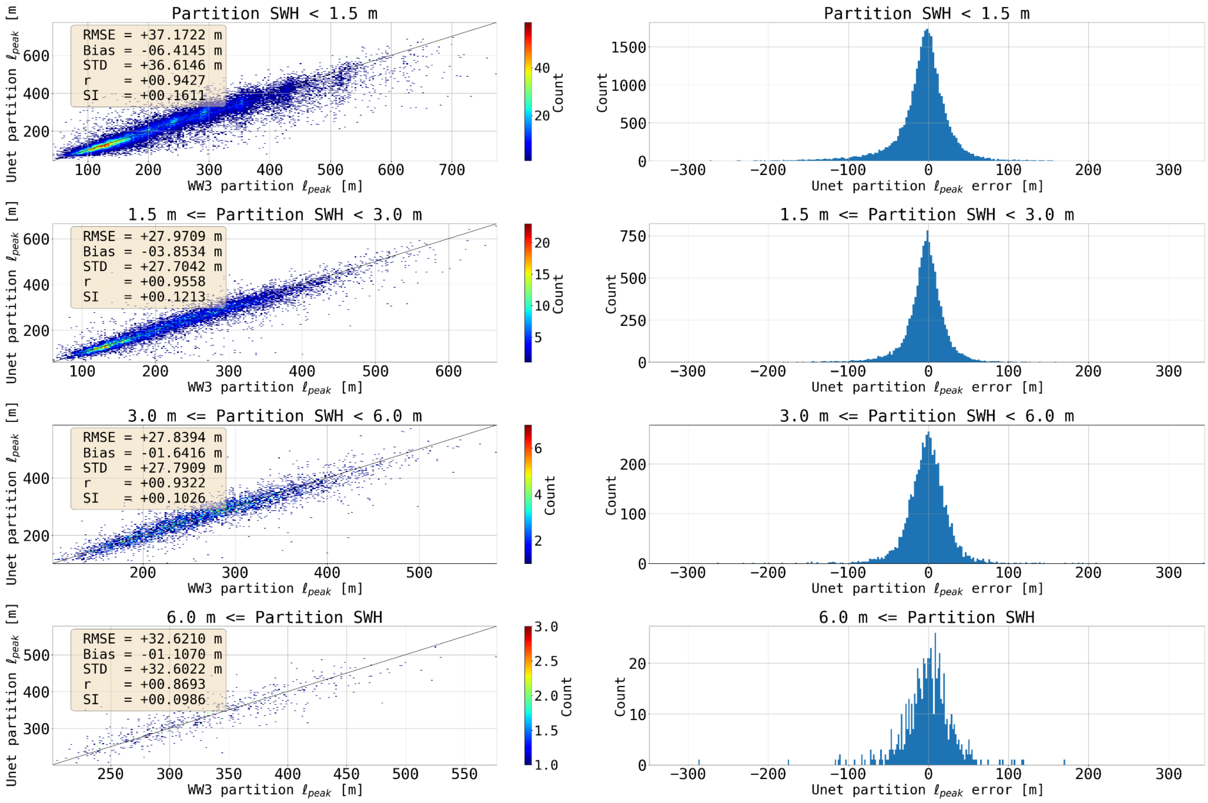

Figure 36: [Left] Comparison of partition peak wavelength for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra partition peak wavelength vs Ground truth colocalized WaveWatch3 directional wave spectra partition peak wavelength. [Right] Partition peak wavelength error histograms for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

Figure 36 presents partition peak wavelength scatterplots and error histograms for four different sea state domains (w.r.t. partition SWH). Interestingly, partition peak wavelength performance peaks for highly energetic sea states (3.0 m < partition SWH < 6.0 m), despite the fact that such sea states only account for around 10% of the dataset. Partition peak wavelength retrieval accuracy subsequently degrades as we move towards both lower energy and higher energy sea states, and reaches its lowest value for low energy sea states (partition SWH < 1.5 m), even though this is the most predominant sea state domain in the dataset (with around 65% of observations belonging to this sea state domain). This indicates that partition peak wavelength retrieval performance is less sensitive to sea state domain observation density than partition SWH retrieval performance, and further reinforces our understanding that performance issues are mostly related to the degree of smoothing of the directional wave spectra produced by the proposed algorithm, which should be improved with a more adequate tuning of network architecture and hyperparameters.

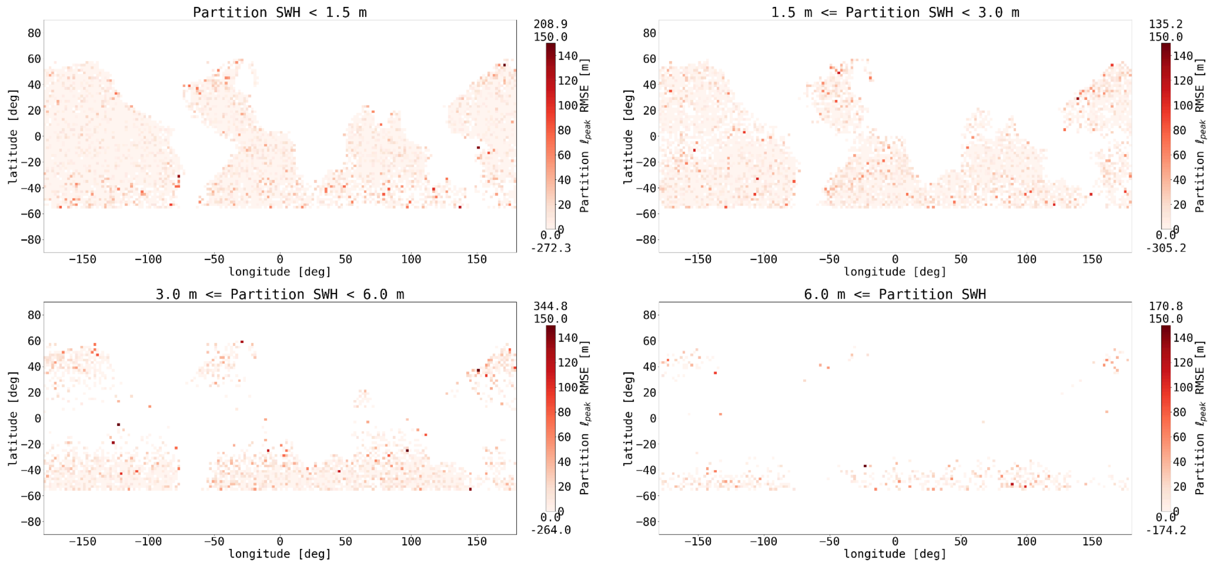

Figure 37: Partition peak wavelength RMSE geographical distribution for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

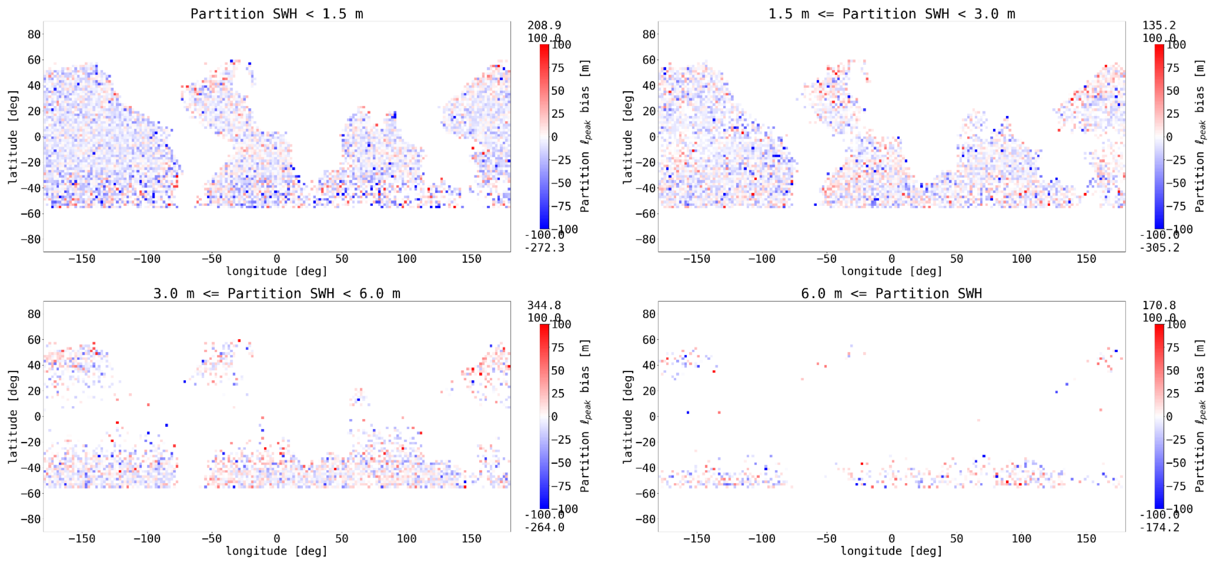

Figure 38: Partition peak wavelength bias geographical distribution for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

Regarding the geographical distribution of partition peak wavelength performance metrics for the considered sea state domains (w.r.t. partition SWH), as depicted in Figures 37 (for peak partition wavelength retrieval RMSE) and 38 (for peak partition wavelength retrieval bias), low energy sea states (partition SWH < 1.5 m) present lower partition peak wavelength performance towards the southern oceans, which is to be expected given the lower observation density for these sea states in this region (as observed in Figure 30). For more energetic sea state domains, no clear geographical patterns in partition peak wavelength performance metrics seems to appear, with homogeneous RMSE and bias distributions that follow observation density, and outliers being distributed homogeneously as well.

Results Summary:

Table 5 : ODL Partition peak wavelength preliminary results summary

Partition SWH							
Wave height domain (m)							
	RMSE	Bias	STD	r	SI	Count	Data percentage
Global	34.26 m	-5.26 m	33.85 m	0.95	15%	50038	100%
0.0 < partition SWH< 1.5	37.17 m	-6.41 m	36.61 m	0.94	16%	32421	64.79%
1.5 < partition SWH< 3.0	27.97 m	-3.85 m	27.70 m	0.96	12%	11949	22.88%
3.0 < partition SWH< 6.0	27.84 m	-1.64 m	27.79 m	0.93	10%	5062	10.12%
6.0 < partition SWH	32.62 m	-1.11 m	32.60 m	0.87	10%	606	1.21%

2.3.5.4 Partition parameters : Peak Direction (θ_{peak})

Global analysis:

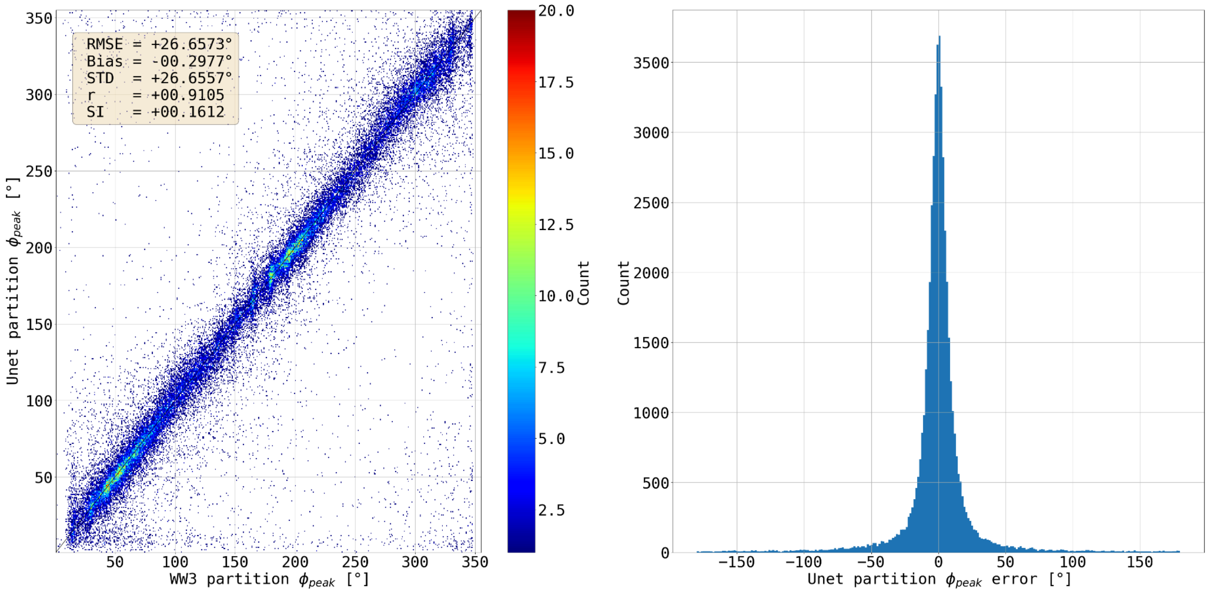

Figure 39: [Left] Comparison of partition peak direction: U-net based directional wave spectra partition peak direction vs Ground truth colocalized WaveWatch3 directional wave spectra partition peak direction. [Right] Partition peak direction error histogram: U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

Figure 39 depicts partition peak direction retrieval scatterplots and error histograms, with a mean RMSE of around 27° and a mean bias of 0.30°. As with partition peak wavelength, current efforts aim at further enhancing retrieval accuracy, which we also hypothesize to be related to output smoothing and may be improved by further tuning network hyperparameters.

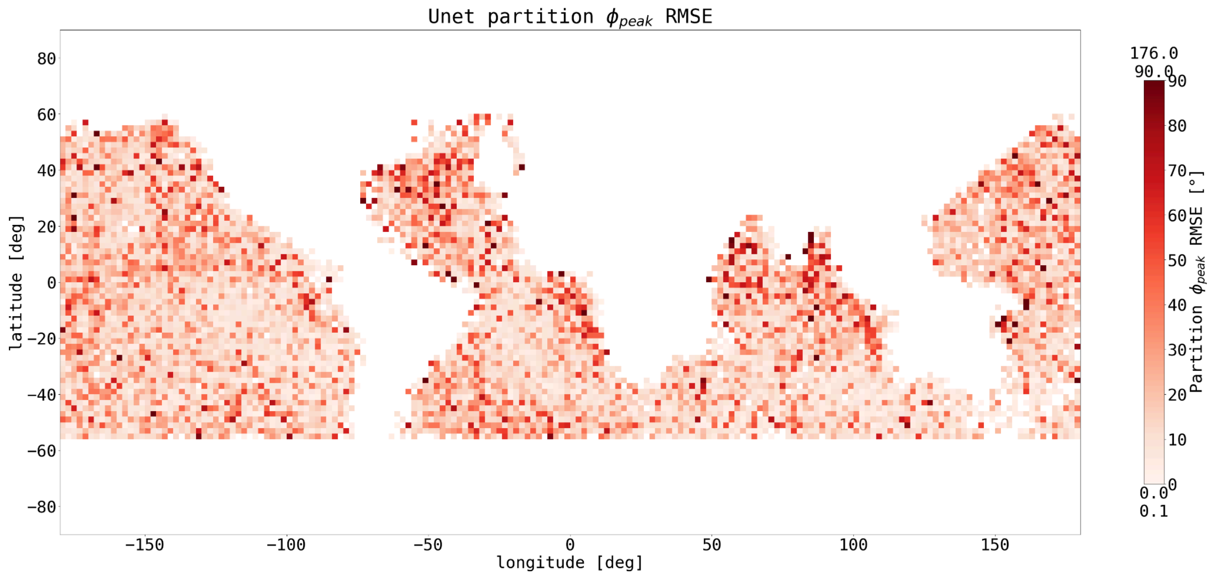

Figure 40: Partition peak direction bias geographical distribution: U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

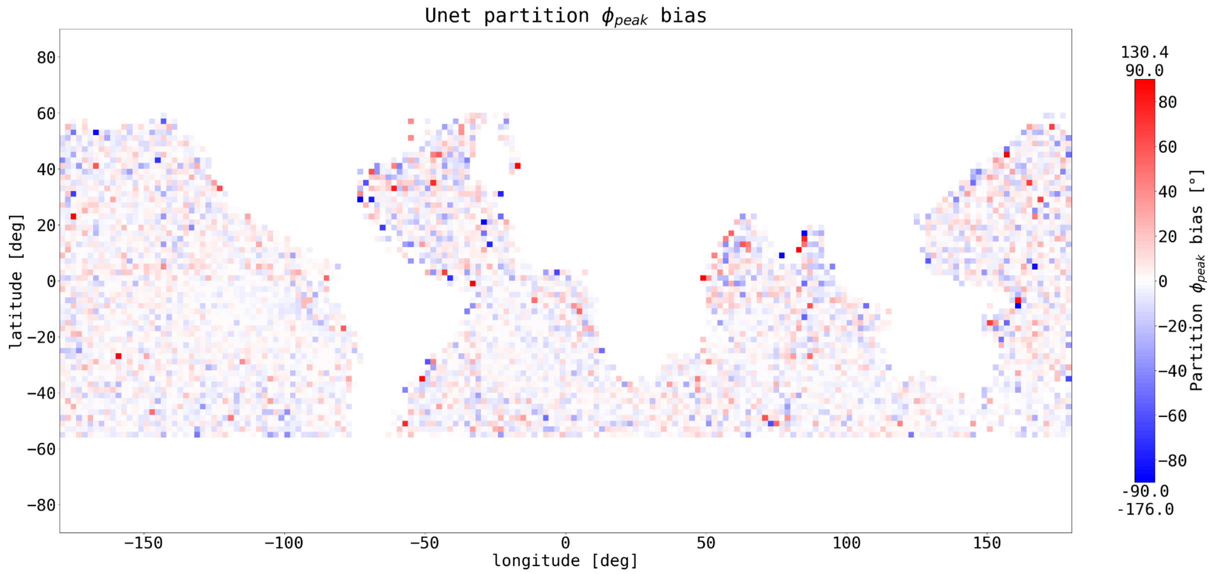

Figure 41: Partition peak direction bias geographical distribution: U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

Regarding the geographical distribution of partition peak direction retrieval RMSE and bias, depicted respectively in Figures 40 and 41, no clear pattern appears to be present, with lower performance metrics distributing homogeneously across the global ocean, with a slight tendency towards higher RMSE and bias values near coastal zones. This is in agreement with our hypothesis that performance issues are mostly related to the degree of smoothing of the directional wave spectra produced by the proposed algorithm, and could thus be enhanced with a more adequate network parametrization and hyperparameter selection.

Sea states domains analysis (with respect to partition SWH):

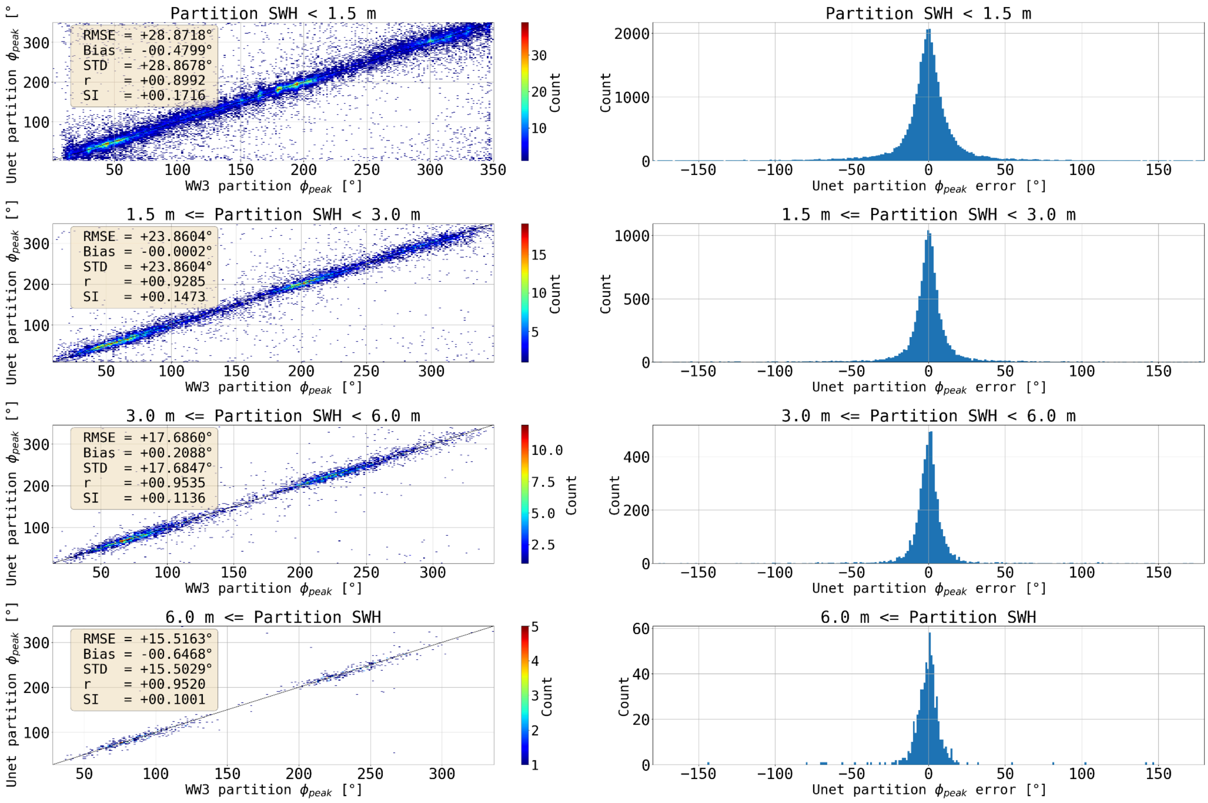

Figure 42: [Left] Comparison of partition peak direction for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra partition peak direction vs Ground truth colocalized WaveWatch3 directional wave spectra partition peak direction. [Right] Partition peak direction error histograms for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

Interestingly, when partition peak direction retrieval metrics are considered for different sea states domains (w.r.t. partition SWH), as depicted in Figure 42, performance seems to improve with sea state energy levels (proxied by partition SWH), in spite of the fact that the number of observations within the dataset becomes increasingly lower for more energetic sea states. Once again, this supports our hypothesis that, similarly to partition peak wavelength, partition peak direction retrieval performance appears to be insensitive to sea state domain observation density, and relates mostly to network architecture and hyperparameter tuning.

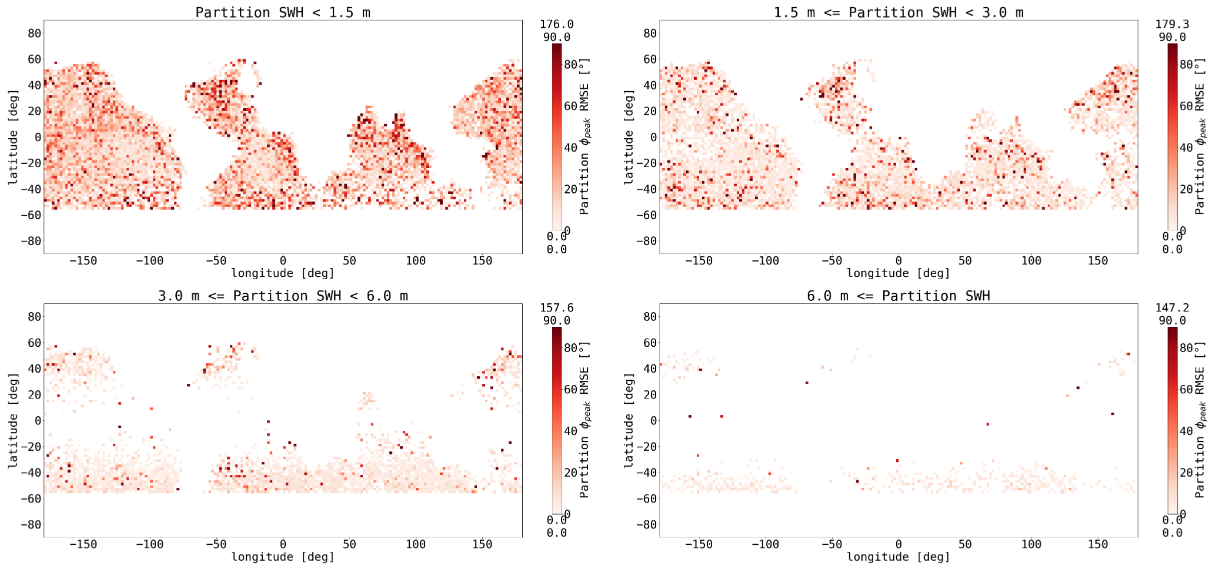

Figure 43: Partition peak direction RMSE geographical distribution for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

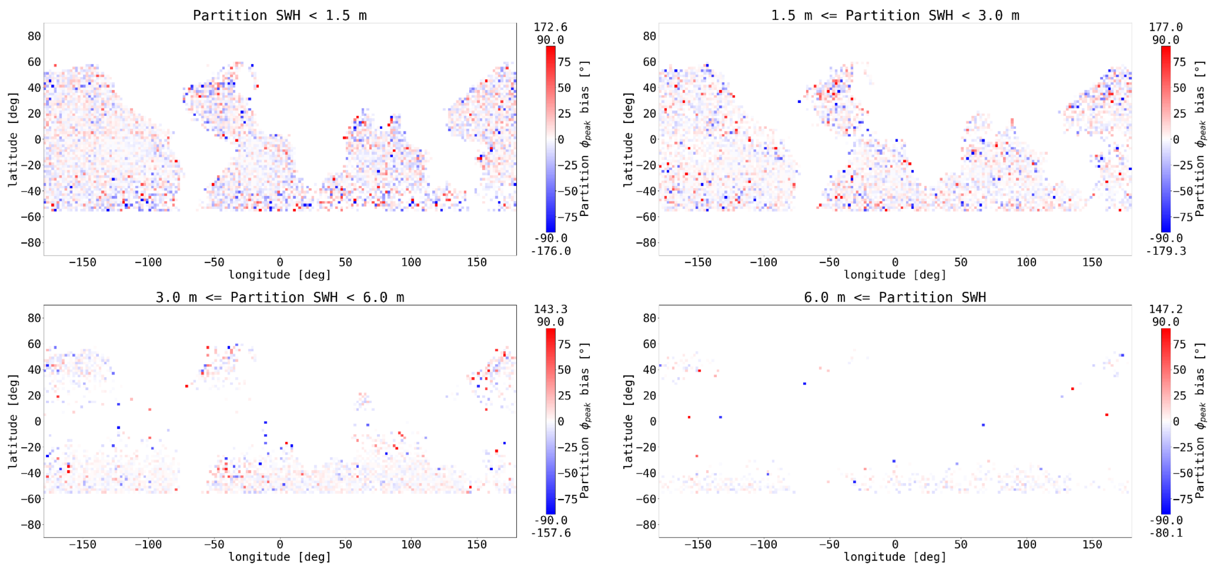

Figure 44: Partition peak direction bias geographical distribution for four different sea state domains (w.r.t. partition SWH): U-net based directional wave spectra vs Ground truth colocalized WaveWatch3 directional wave spectra.

When considering the geographical distribution of partition peak direction RMSE and bias, depicted in Figures 43 and 44 respectively, RMSE and bias appear, globally, to follow the geographical observation density distribution for different sea states domains depicted in Figure 30, particularly for lower energy sea states, whereas higher RMSE and bias values, and outliers, seem to be marginally more predominant closer to coastal zones.

Results Summary:

Table 6 : ODL Partition peak direction preliminary results summary

Partition SWH							
Wave height domain (m)							
	RMSE	Bias	STD	r	SI	Count	Data percentage
Global	26.66°	-0.30°	26.66°	0.91	16%	50038	100%
0.0 < partition SWH< 1.5	28.87°	-0.48°	28.87°	0.90	17%	32421	64.79%
1.5 < partition SWH< 3.0	23.86°	-0.00°	23.86°	0.93	15%	11949	22.88%
3.0 < partition SWH< 6.0	17.69°	+0.21°	17.68°	0.95	11%	5062	10.12%
6.0 < partition SWH	15.52°	-0.65°	15.50°	0.95	10%	606	1.21%

2.3.6 Specificity of ERS-2 gyroless period

For the proposed ERS-2 gyroless option, we will evaluate the effect of the different ERS-2 gyro-less piloting modes on partition parameter retrieval performance and uncertainty. The proposed approach relies on:

1) Applying the ERS-2 base neural network on ERS-2 cross-spectra from the gyro-less period to produce a baseline inversion of wave-spectra for the different piloting modes
2) Evaluating partition parameter retrieval performance and uncertainty on this baseline, for each piloting mode, w.r.t colocated WaveWatch-III wave spectra
3) Identifying suitable improvement strategies to the base approach to mitigate the effects on partition retrieval performance of the different gyo-less operation piloting modes
4) Re-applying the improved neural network to ERS-2 cross-spectra for the considered gyro-less period
5) Evaluating partition parameter retrieval performance and uncertainty on the corrected neural network inversion
6) Quantifying improvements in partition parameter retrieval performance and uncertainty w.r.t. the baseline neural network inversion retrieval.

Partition parameter retrieval uncertainty characterization for ERS-2 Gyro-less will follow the same design logic and structural principles applied for Sentinel-1 WV, as described in Section 2.3.4.

At the current stage, a colocalized dataset of ERS-2 cross-spectra and WaveWatch-3 model wave spectra for the whole mission archive is under preparation. This dataset will serve as the basis for training both the base ERS-2 model and deploying the ERS-2 gyro-less evaluation and model adaptation strategy detailed above. The evaluation of uncertainty and error propagation for ERS-2 gyro-less will also follow a two tier approach, and will be performed both before and after corrections to the base ERS-2 neural network on a hold-out dataset not exploited for model training/adaptation, to ensure comparison accuracy and prevent cross-contamination between data used for training/improving the model and data used to validate the approach. This analysis will provide important information for quantifying improvements on partition parameter retrieval uncertainty, and thus provide relevant metrics for quantifying and validating the identified improvements and model adaptation strategies in terms of their efficacy for mitigating the effects of ERS-2 gyro-less operation on partition parameter retrieval performance.

Mirroring the approach used for Sentinel-1 WV data, preliminary internal verification results will provide a first estimate of the algorithm's performance. Internal verification statistics will be computed per partition parameter, and a detailed geographical and sea-state based analysis will be performed, mirroring the preliminary results analysis presented for Sentinel-1 WV presented in Section 2.3.1.1. Similarly to Sentinel-1, internal verification statistics for ERS-2 can be considered as an adequate proxy for uncertainty at an initial stage, provided that they are computed on ground truth data from the same source used as groundtruth during training (in this case, numerical model outputs from WaveWatch III). A second-stage thorough evaluation of uncertainty and error propagation is to be performed once the model adaptation approach has reached sufficient maturity, for both internal verification and validation purposes, before the reprocessed ERS-2 gyro-less data can be added into the CCI+ Sea State data record.

2.3.7 Specificity of CFOSAT Off-Nadir (Option)

The proposed CFOSAT Off-Nadir option involves two different tasks, namely the derivation of a quality flag for SWIM-derived partition parameters, consistent with current Sentinel-1 partition parameters and associated quality indicators, followed by the derivation of an intercalibration approach between Sentinel-1 and SWIM derived partition parameters.

Both approaches will rely on a high confidence dataset of cross-assigned partitions and associated partition parameters from SWIM, Sentinel-1, numerical model reference data and independent SOFAR Spotter drifter measurements, to ensure the quality flagging and intercalibration approaches are derived from physically and algorithmically meaningful correspondences.

The proposed approaches explicitly account for full wave spectra inversion stability, full wave spectra quality evaluation and validity, partition cross-assignment uncertainty and mismatch, and partition parameter retrieval errors, uncertainty and differences between missions. In this regard, this multi-level approach separates spectral inversion uncertainty, partition cross-assignment uncertainty and partition parameter uncertainty and inter-mission differences by design.

2.3.7.1 - Uncertainty estimation for the development of a SWIM partition parameter quality flag

Initially, a quality flag for SWIM partition parameters will be derived by estimating SWIM partition parameter retrieval uncertainty from partition parameter retrieval residuals w.r.t both model (WaveWatch3) and in-situ (SOFAR Spotter drifters) reference data, as a function of internal predictors computed from SWIM-derived wave spectra partitions. The derivation of the uncertainty estimates and the quality flag will exploit a multi-level approach that explicitly and independently accounts for uncertainty sources at the whole wave spectra, partition cross-assignment and partition levels.

Importantly, the uncertainty estimates from which the quality flag will be derived will be computed from partition parameter retrieval residuals with respect to both colocalised model reference data as well as independent in-situ measurements, thus exploiting multiple independent reference data sources for uncertainty characterization. In this regard, the derivation of the quality flag methodology for SWIM partition parameters relies on partition parameter retrieval residuals, w.r.t both numerical models and in situ data, as a proxy for uncertainty characterization, which is aligned with the uncertainty characterization approach presented in Section 2.3.1.1 for Sentinel-1 WV partition parameters retrieval.

As such, the computation of the quality flag will be based on a first estimate of SWIM partition parameter retrieval uncertainty, which will complement the current SWIM L2S product and improve on existing SWIM L2S partition parameter uncertainty estimates and quality information, currently limited to the presence of land or shallow water or missing pulses within the footprint used to estimate the spectrum in the radar look direction, and lacking quality indicators related to observation conditions, such as low wind or presence of non-wave related inhomogeneities that could create erroneous, exaggerated or artificial swell partitions.

2.3.7.1 - Uncertainty evaluation for intercalibration

Subsequently, a subset of high quality correspondences between SWIM and Sentinel-1 derived wave spectra partitions will be exploited to develop an intercalibration approach between SWIM and Sentinel-1 derived partition parameters. Similarly to the quality flag development framework, the proposed intercalibration methodology will rely on a multi-level approach to explicitly account for and separate spectral inversion uncertainty, partition cross-assignment uncertainty and partition parameter retrieval uncertainty, and ensure the intercalibration is derived from an ensemble of high quality, geophysically meaningful partition correspondences and associated partition parameters.

The intercalibration relies on exploiting cross-assigned SWIM, Sentinel-1 and WaveWatch-3 wave spectra partitions and associated partition parameters to perform an initial relative intercalibration between Sentinel-1 and SWIM partition parameters, as well as colocated SOFAR Spotter drifter measurements to further correct the initial SWIM/Sentinel-1 relative intercalibration w.r.t. independent in-situ reference data. Importantly, since wave spectra inversion approaches for both SWIM and Sentinel-1 are derived using WaveWatch-3 wave spectra as reference during training, exploiting SOFAR Spotter drifters in situ measurements to provide an additional correction after relative intercalibration is expected to further mitigate the effects of any remaining common model biases shared between missions.

The evaluation of intercalibration performance will rely on residual statistics between SWIM and Sentinel-1 derived partition parameters, as well as partition parameter retrieval residuals relative to both WaveWatch-3 numerical wave model outputs and SOFAR Spotter drifters in-situ observations, computed on an independent hold out dataset excluded from both uncertainty modelling and calibration coefficient estimation to avoid information leakage, and analysed both globally and per sea-state class.

In this regard, the evaluation will rely on the preliminary uncertainty analysis presented in Section 2.3.1.1 for Sentinel-1 WV partition parameters as an initial comparison benchmark. The same analysis will then be carried out to characterize Sentinel-1 WV partition parameter retrieval uncertainty after intercalibration, and evaluate performance gains provided by the intercalibration approach in terms of partition parameter retrieval uncertainty. Importantly, following the approach presented in Section 2.3.1.1, this characterization will rely on internal verification statistics w.r.t. reference data from both WaveWatch-3 numerical wave model outputs and in-situ SOFAR Spotter drifters measurements as an uncertainty proxy to evaluate the effects of the developed intercalibration approach on partition parameter retrieval uncertainty, both globally and per sea-state, and characterize and quantify any remaining global or sea-state dependent biases that may have not been corrected by the intercalibration approach.

2.4 Altimeter Sigma0

Normalised backscatter i.e. sigma0 is the parameter that is formally being evaluated within the sea state CCI project, but the implications for wind speed products are described later in this document, as they will be generated under one of the CCI options. There are a number of sources of error in sigma0, ranging from random ("noise") terms associated with particular realizations of the sampling of the sea surface, to errors in the retrieval algorithms and errors in the corrections. These are discussed briefly here.

2.4.1 Algorithm

A number of different approaches may be used to estimate the total power represented by the return waveform. One approach would be to sum the power in all the individual wavebins, but that would be sensitive to variations in positioning of the waveform within the window. Therefore it is much more common to fit a model shape to the observed power levels and derive the amplitude of that model fit. The most commonly used models are MLE-3 and MLE-4 (Maximum Likelihood Estimator fitting either 3 or 4 parameters), and WHALES is very similar.
The 4th parameter in MLE-4 corresponds to "platform mispointing" but in most cases it represents an overfit to the data causing spurious variations in the σ^0 estimate. Figure 45 shows the variability in the 20 σ^0 estimates (from overlapping footprints) that contribute to each 1 Hz mean value. In a manner analogous to that shown for Hs (Fig. 4) we can calculate the mean measure of variability associated with given mean conditions. It is readily apparent in Fig. 45b that the estimates using MLE-4 are much noisier than for MLE-3.

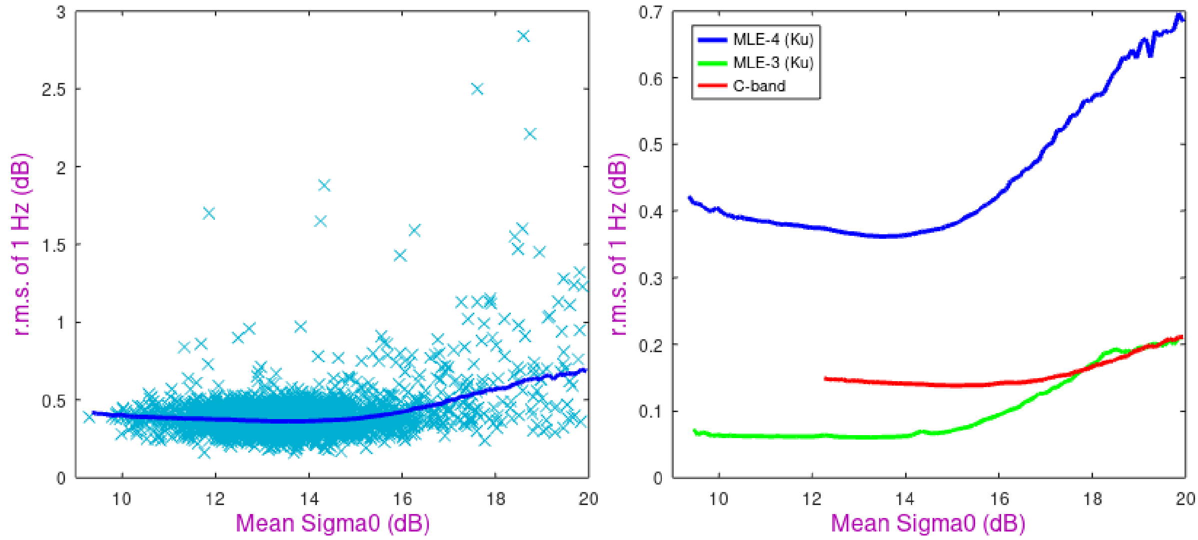

Figure 45 : Estimate of the intrinsic noise within altimeter estimates of sigma0. Left panel shows r.m.s. of the 20 MLE-4 Ku-band estimates contributing to every 1 Hz mean value, with the solid blue line showing the mean in each narrow σ^0 bin. Right-hand side compares the results for MLE-3 and MLE-4 for Ku-band and C-band. (All data from Jason-2.)

Work by Quartly (2009) showed that the r.m.s. difference between the Jason-1 and Jason-2 instruments during their tandem phase was also reduced by a factor of 3 by replacing the MLE-4 estimates with "adjusted values" (similar to MLE-3).

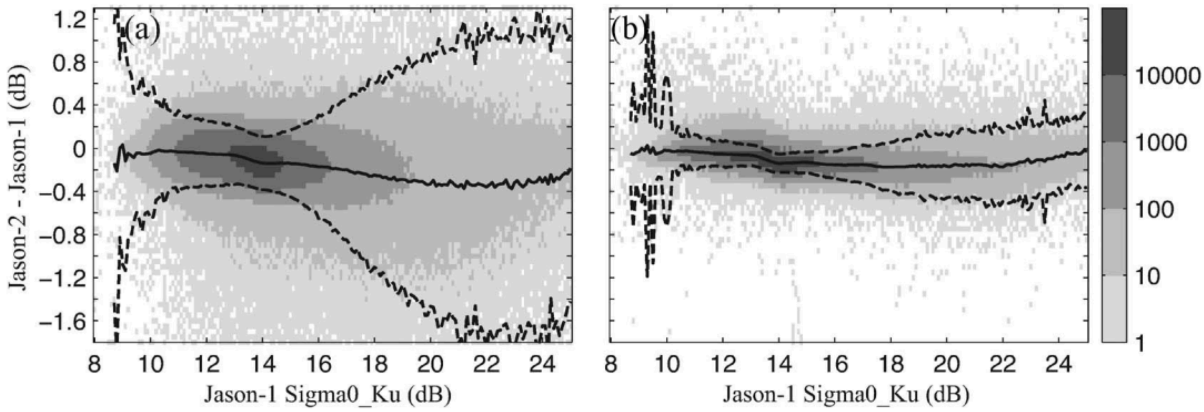

Figure 46 : 2-D histograms of the σ^0 difference between Jason-1 and Jason-2 at Ku-band, with the added line showing the mean for each narrow σ^0 bin plus the spread (indicated by ±2 std. dev.). a) MLE-4 estimates b) Adjusted estimates. [Taken from Quartly (2009).]

2.4.2 Instrument corrections

The Automatic Gain Control is a system of imposing different levels of attenuation in the receiver chain. However, the amount of attenuation applied may not be correctly known, so there can be non-linear errors associated with the AGC settings. This is particularly the case for Jason-2, where the discrepancy between nominal AGC setting and the actual attenuation can be up to 0.07 dB. However, if evaluated and corrected for, the estimated accuracy of these implemented attenuators is better than 0.01 dB (see Fig. 47).

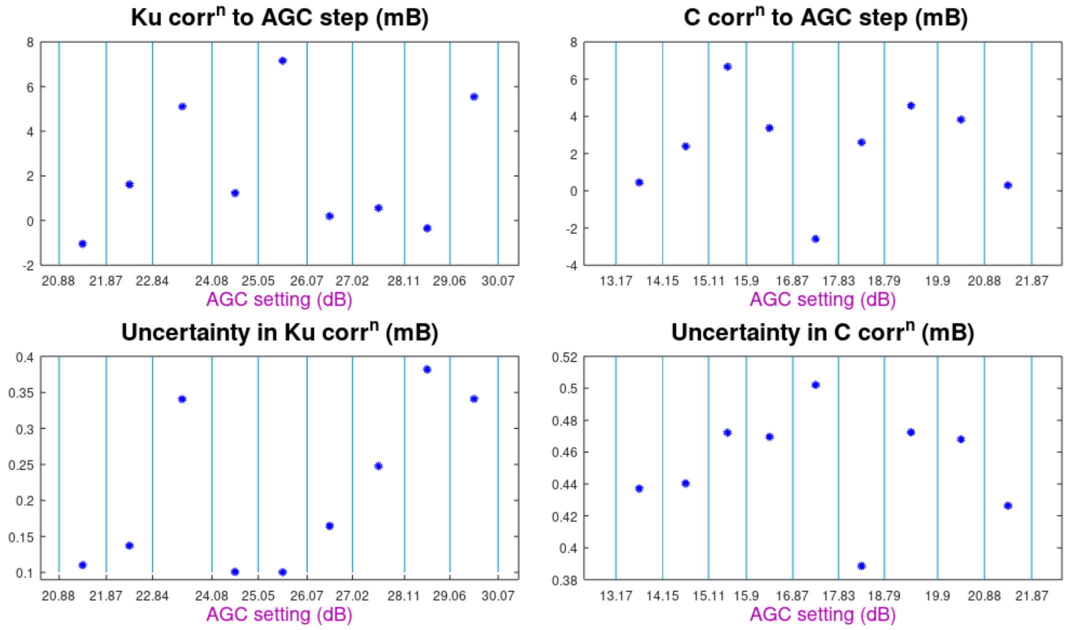

Figure 47 : Top row gives the mean error (bias) in transitions between successive AGC settings for Jason-2. Bottom row shows the uncertainty in those biases (1mB = 0.01 dB.) [Taken from Quartly (2025).]

Another aspect is the estimated correction for attenuation by the atmosphere. A synoptic estimate is derived from the observed brightness temperatures of the microwave radiometer (MWR) that is on the spacecraft aligned with the altimeter boresight. This may have problems in the coastal zone, when land enters the radiometer footprints. Alternatively an independent estimate of the attenuation can be provided by a numerical weather prediction model. Fig. 48 shows that the mean bias between these two is rarely more than 1 mB, with a std. dev. of ~2 mB

.

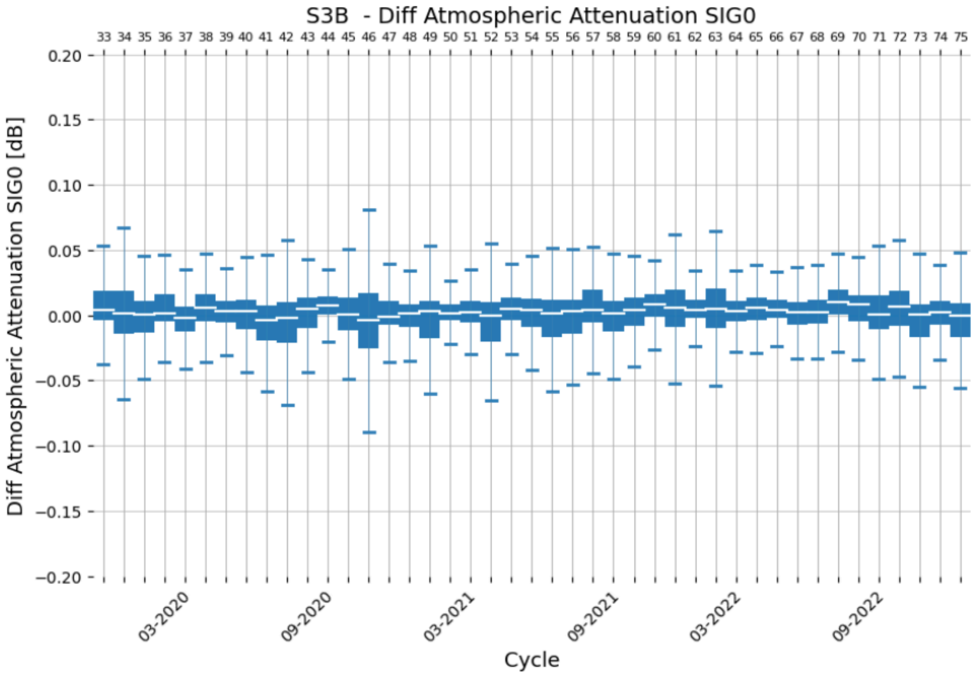

Figure 48 : Time series of difference MWR-ECMWF of Atmospheric attenuation of Sigma0 (SAR) for Sentinel-3B. [Fig. 28b from MWR cyclic performance report, produced by S3MPC-STM.]

2.4.3 Uncertainty in Wind Speed

There are two main sources of uncertainty in derived wind speed. Firstly there is the uncertainty in the estimated sigma0 value. This corresponds to both a measurement error, including how well the waveform conforms to the retracker model (whether this be MLE3 or some numerical scheme) and the error in the sigma0 adjustment of that particular sensor/epoch. The work on the sigma0 homogenization is anticipated to produce sigma0 values that are consistent to within 0.03 dB or better, but there may be a small proportion of non-Brown-like waveforms that escape the flagging procedures, with these individual 1 Hz records having greater errors. Typically an uncertainty of 0.03 dB translates to a wind speed error of 0.1 m s^{-1}.

The second source of error is the choice of a wind speed algorithm. This is to be addressed later in the CCI project as one of the selected options. For now, an indication of the likely error can be ascertained by comparing the output of a number of commonly used wind speed algorithms. Figure 48bis compares the output of the Abdalla, (2012), Gourrion et al. (2002) and Collard (2005) algorithms, noting that the latter two are also functions of wave height. In this example the series of input sigma0 values to an algorithm have been bias adjusted to give generally good agreement with the other algorithms at moderate wind speeds. In this comparison one notes:

i) The Gourrion algorithm shows much stronger dependence on wave height than the later Collard algorithm (with Abdalla having no wave height dependence).

ii) The extrapolation of Gourrion for low sigma0 values is based on observations by Young (1993) and differs from the other estimates Indeed there are fewer in situ match-ups to evaluate inversions in high wind speed conditions, so algorithm-based uncertainty is likely to be greater.

iii) It has been noted that for wind speeds below 2 m s^{-1}, the small scale ripple structure is affected more by other factors than wind speed. The Abdalla algorithm has a much higher asymptote than the others. Although algorithm-based errors could be as high as 50% in this region, for almost all applications it is sufficient simply to know that wind speed is below 2 m s^{-1}.

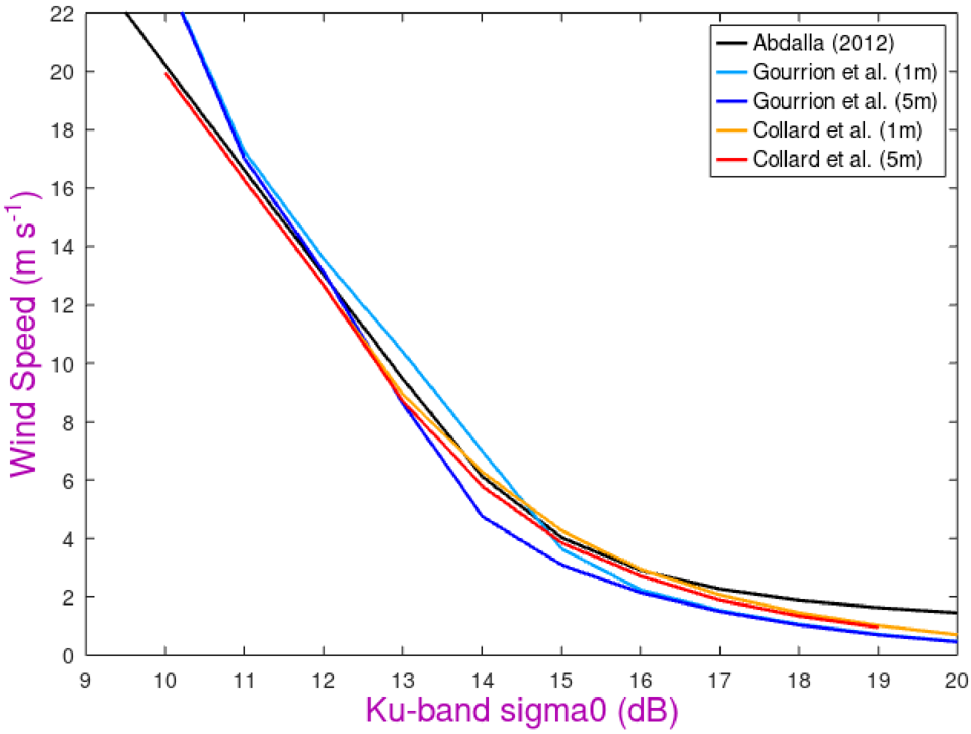

Figure 48bis : Comparison of 3 wind speed algorithms. All are based on Ku-band σ^0 values, which have been bias corrected to give broad agreement. The Gourrion and Collard algorithms also incorporate Hs values in producing their estimates.

2.4.3 Long-term Trends in Missions and Inter-mission Biases

Estimating and correcting for the long-term changes both within missions and between different missions is the key focus of this CCI work. The dual-frequency technique was able to detect a step change in S6-A processing, estimating the change to within 0.01 dB. It is expected that any gradual deterioration of the instrument can be monitored to that accuracy provided that it is constant over the entire dynamic range used over the ocean. Extraneous sources of noise unrelated to the signal could be harder to correct for.

The analysis technique depends upon the constancy of the relationship between normalised backscatter at the two frequencies at nadir-viewing. The mispointing problems with Jason-1 challenge this assumption, and no estimate has yet been made of the uncertainty due to this issue. It also follows that the technique is only applicable to compare two (P)LRM missions as σ^0 values associated with SAR waveforms have a different Ku-C relationship. An estimate of the error associated with aligning different missions will be determined later in the project through dual satellite crossovers, but these might be limited because a short time separation is needed because of the changeable nature of winds.

2.5 TerraSAR - X (Option)

TerraSAR-X (TS-X) StripMap (SM) scenes (ca. 30×50-70 km, pixel spacing 2.5-4.5 m) were validated with MFWAM reanalysis results. However, as the model can not reproduce direct-to-coast effects, validations are performed only for collocation >10 km from the coast. As the TS-X are dominantly acquired over coasts and harbors, only cay. 30% of all processed TS-X data are included into validation analysis. Figure 49 presents validations for TS-X/TD-X SM for significant wave height Hs and Tm2 wave period (all data > 10 km from the coastal line). The 30×50 km TS-X/TD-X SM scenes are processed with a raster of 1.5 km (ca. 20×30=600 values per a 100% ocean scene) and present the local sea state distribution by comparison to coarse model grid points 1/5°.

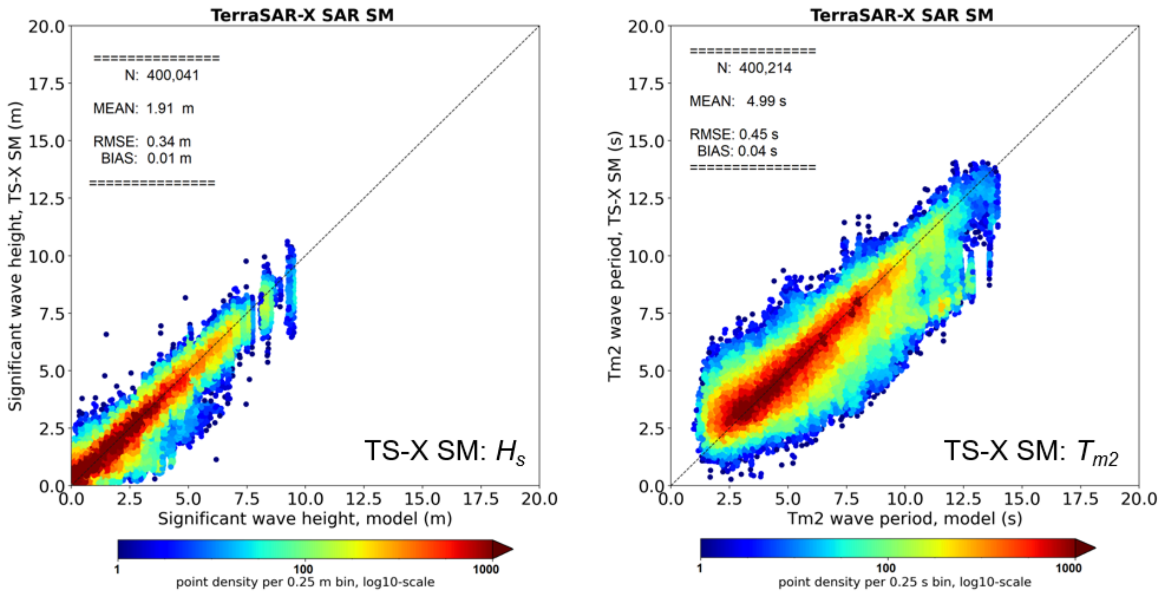

Figure 49: Validations for TS-X/TD-X SM for significant wave height H_s and T_{m2} wave period for data > 10 km from the coastal line. The 30×50 km scenes are processed with a raster of 1.5 km and present the local sea state distribution by comparison to coarse model grid points 1/5°.

The RMSE distribution for Hs and for different sea state domains (partition) is presented in Table 8. As can be seen, most of the data are acquired in coastal areas with Hs under ca. 5 m (ca. 90%, mean Hs~1.9 m), only individual acquisitions include strong sea states with Hs higher than 6 m.

Table.8: TS-X/TD-X SAR SM RMSE distribution for Hs and for four sea state domains

Sea state domain, Hs (m)	fraction (%)	RMSE (m)
0.0 –1.5	58	0.31
1.5 – 3.0	34	0.37
3.0 – 6.0	6	0.55
6.0 <	2	0.67
total	100	0.34

A cross-comparison for collocated S1 IW, TS-X and buoys were carried out and shown in Figure 50. The cross-comparison matrix presents RMSEs for each combination for all four sources including the model. Unfortunately, such quadruple collocations are very rare (N=152 means number of all S1-IW subscenes collocated to TS-X subscenes and also collocated to a buoy). Statistically it can be only summarised that RMSE values are on the same order as the uncertainties for each equipment.

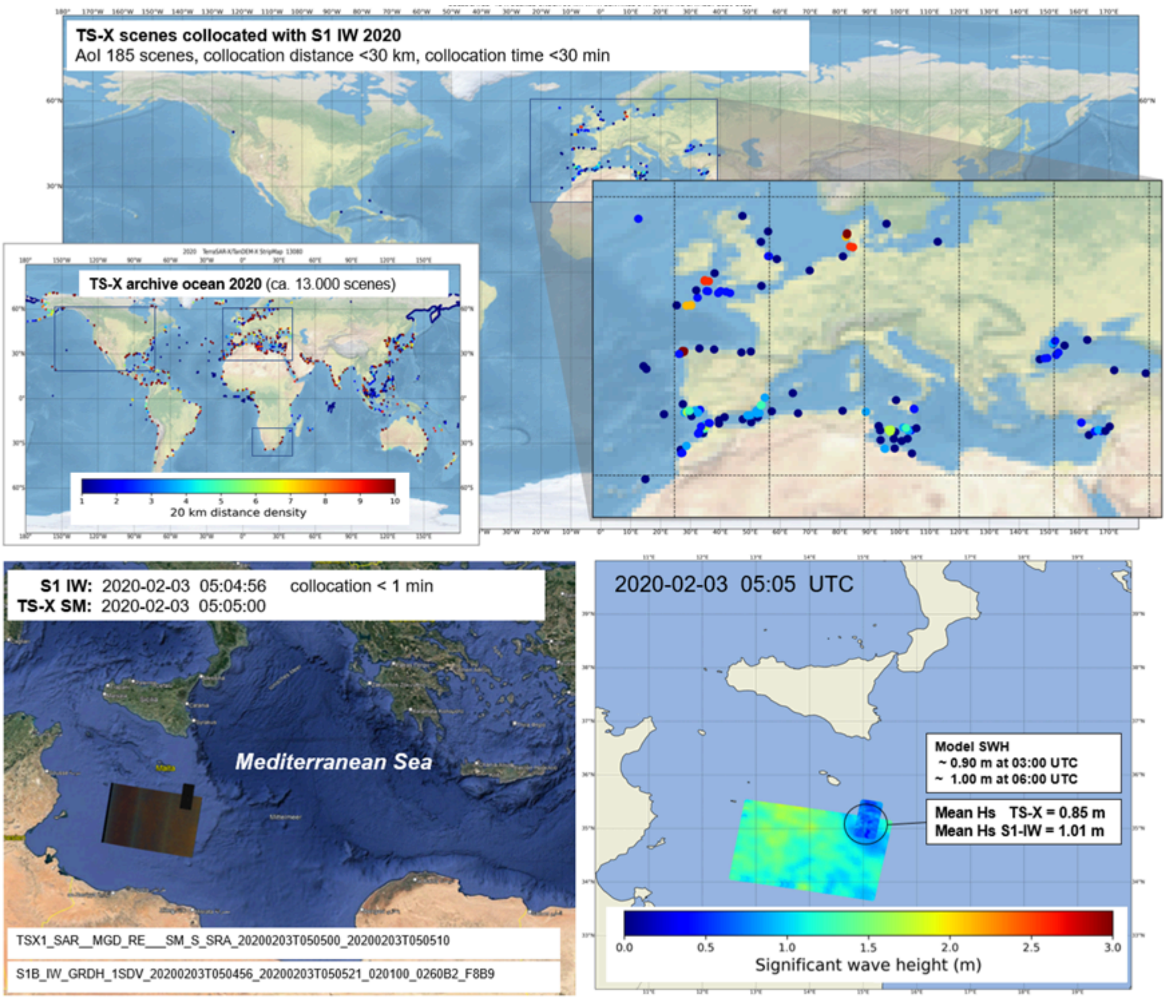

Figure 50: TS-X/TD-X SM scenes from the ocean archive 2020 collocated with S1 IW scenes in Europe and North America with 30 km and 30 min (first row, the whole ocean archive displayed bottom left), the colour means scenes density within 20 km considering

TS-X SM scene centres. An example of a direct collocation under 1 min between S1 IW and TS-X SM (second row).

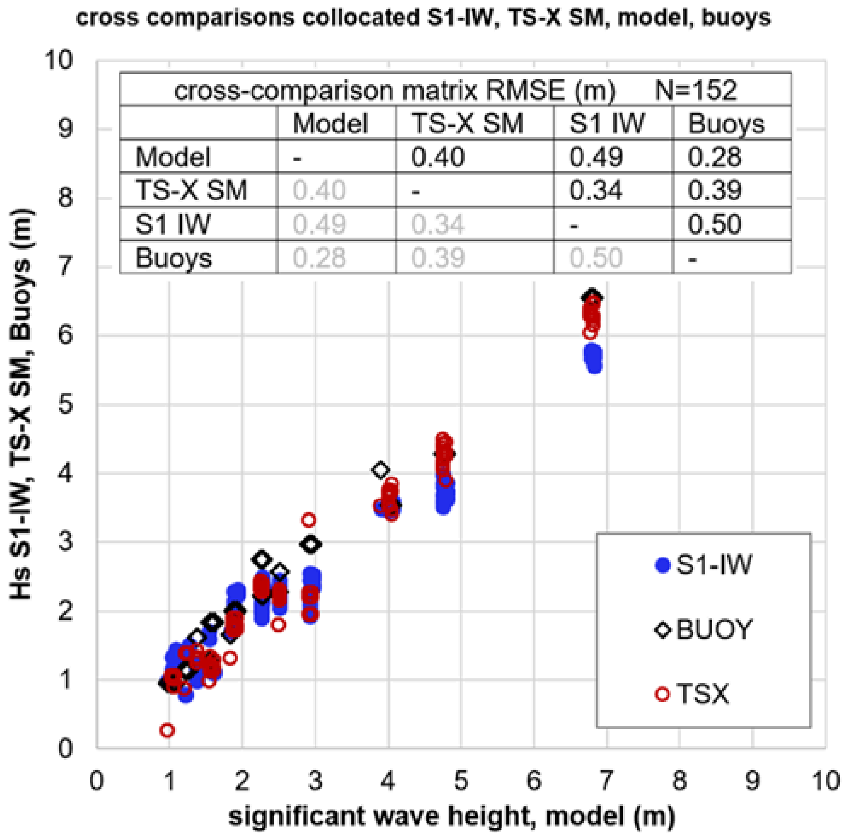

Figure 51: Cross-comparison for quadruple collocations S1 IW, TS-X, buoys and model for 2020 and 2021 archived data. The RMSE cross-comparison matrix for all four sources. The partial RMSEs are calculated only for locations where all four sources are available.

3. Error Propagation

3.1 Altimeter Hs

There are two main types of error contributing to the overall error for altimeter data: that due to fading noise, which will be independent from one waveform to another, and that due to systematic errors (incorrect assumptions in algorithm or effect due to swell or wave direction), which are likely to be consistent over large scales (up to 100 km). These can be assessed by comparison with a large number of in situ observations, but these need to span a wide variety of conditions (wave height, swell period, wind conditions), and in turn be fully and reliably calibrated. It will be particularly challenging to assess any systematic errors at very high seas states. The overall error is given by:

$$(\text{Overall error})^2 = (\text{Fading error})^2 / \text{No. of waveforms} + (\text{Systematic error})^2 \quad (2)$$

This is illustrated in Fig. 49, for two different values of fading error (representing what is currently achieved, and what could be with improved algorithms) and two levels of systematic noise. For comparisons with buoys and models, it is more important to reduce the systematic error; reducing sensitivity to fading error is mainly beneficial for fine resolution studies e.g. near the coast or in response to well-defined current features.

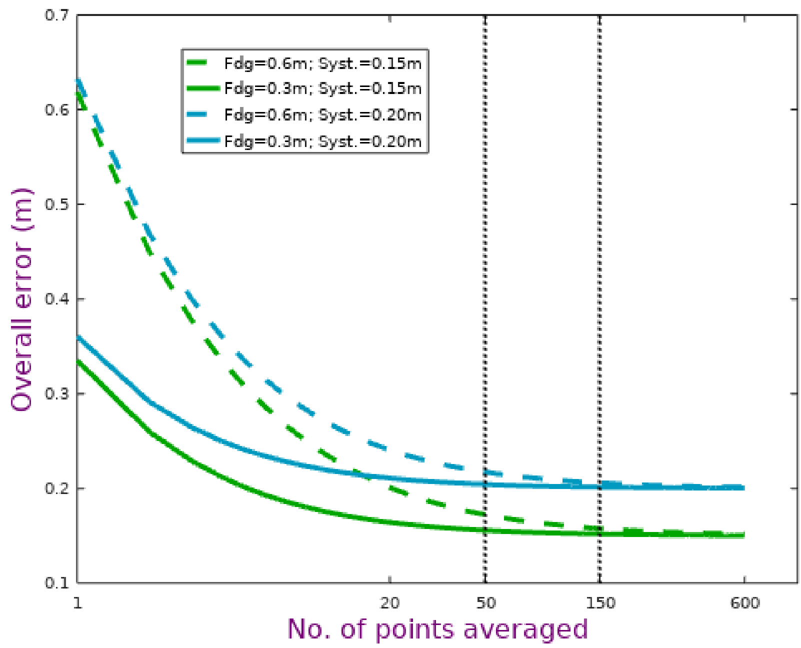

Figure 49 : Illustration to show that error in altimeter estimates due to fading noise dominates for scales of 1 Hz (20 obs.) and finer, but systematic errors are more important at larger scales. The dashed line at 50 pts represents the averaging scale used in buoy comparisons, and that at 150 pts equates to 50 km, a typical scale of high-resolution wave models.

In the production of L4 gridded products (see Section 4) there may be some further reduction in uncertainty due to averaging of observations from different days; however regions with persistent existence of swell or dominant wave directions, will suffer from geographically-correlated errors.

3.2 Altimeter Sigma0

Ultimately, the quality of the CCI dataset σ^0 dataset will be assessed via wind speed. There are many different wind speed algorithms, with Fig. 50 comparing a few of them. The Abdalla (2012) algorithm is the simplest of the three, since it does not involve Hs. For σ^0_{Ku}<14 dB the slope is roughly -3m s^{-1} dB^{-1}, so an error of 0.1 dB in σ^0 leads to a bias of -0.3 m s^{-1} in the wind speed estimate. Fig. ZZ1 also showed the effect that a different choice of wind speed inversion algorithm could have, with the greatest uncertainties being for the extrapolation to high wind speeds. The choice of one over another will lead to a systematic error in winds. As some proposed algorithms use Hs as a second input parameter; errors in Hs can also feed into the uncertainty of the wind speed. The ongoing work to develop a wind speed algorithm will investigate this, but from the example curves for Gourrion et al and

Collard et al. in Fig. ZZ1, it can be seen that an error in Hs of 0.5 m will only have a small impact on derived wind speed.

For the L4 monthly gridded products, there is also uncertainty in the mean associated with the temporal sampling. When producing a climatology on a 2°x2° grid, a satellite in the TOPEX/Jason orbit will typically have 6 independent passes through each square (see Figure 50). (It will be very similar for all other altimeter missions.) The standard deviation in each square indicates the error in assessing the mean from one transit. The uncertainty in a single transit appears to be ~2 m s^{-1} for environments where the mean wind speed is less than 8 m s^{-1} and then rising with mean conditions to 3.5 m s^{-1}. (Note there could be some regional variation in the shape and width of wind speed p.d.f.s, but this has not been investigated here. A linear approximation is given by the black dashed line, which predicts that the uncertainty due to sampling in a monthly mean gridded product will be given by:

uncertainty (in m s^{-1}) = [1.25 + 0.16 mean(Hs)] / sqrt(N)

where Hs is wave height in metres, and N = no. of independent transits by all altimeters. (Note, 2 altimeters operating in a tandem mission will **not** be independent.) With 3 independent altimeters operating N will be 18, giving uncertainties in gridded means ranging from 0.35 to 0.8 m s^{-1}.

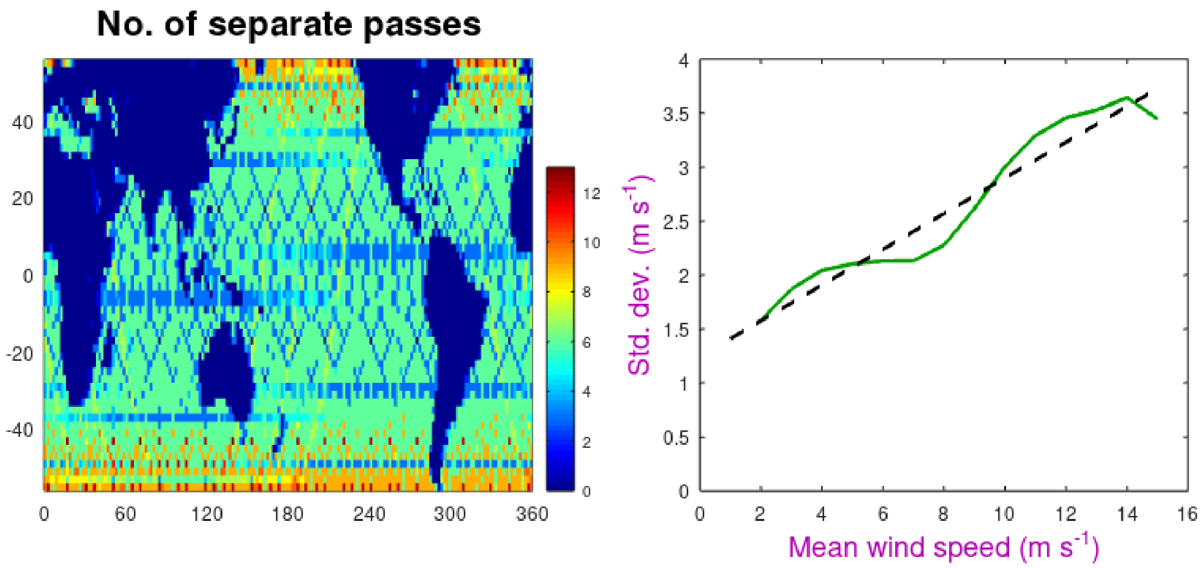

Figure 50: Left-hand panel shows number of Jason-3 passes through each 2°x2° square during a 30-day period. The green line in the right-hand panel shows the standard deviation as a function of mean wind speed, calculated for all those locations with exactly 6 transits (which will correspond to 3 ascending and 3 descending passes in Jason-3's 9.92-day repeating orbit). The black dashed line is a linear fit to the green line.

4. Uncertainty in L4 Gridded Products

4.1 Uncertainty in the Mean

4.1.1 Wave Height

Ultimately the aim of an L4 gridded product is to characterise the distribution of values to be found in a given box over a certain period of time: these are often reduced into measures of mean and maximum. However, the set of observations is far from complete, neither covering all locations in a box or all times, and thus the mean of the altimeter observations will not be the same as the mean of the underlying conditions. This examination is performed for 2.5˚ x 2.5˚ boxes (as this is close to the longitudinal spacing of Jason tracks) for a monthly period (see Fig. 51).

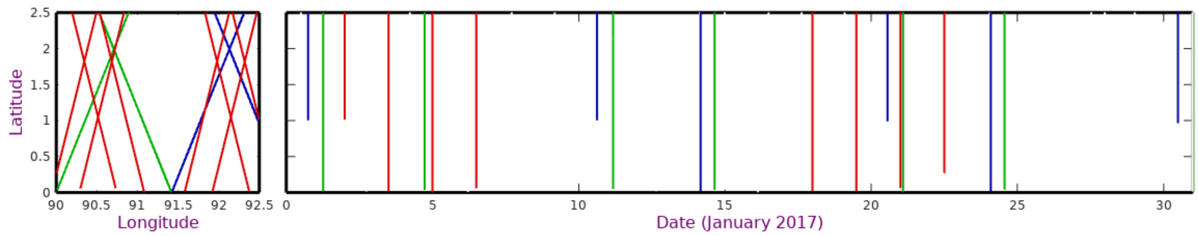

Figure 51 : a) Left panel shows altimeter tracks across a randomly-selected open ocean box for January 2017. Jason-3 (in the reference orbit) is in blue; Jason-2 (in the interleaved orbit) is in green; AltiKa (in its drifting orbit) is in red. [Note AltiKa's track spacing is much finer than when in 35-day repeat, but now it takes twice as long to give roughly uniform coverage.] b) Right panel shows the timing of those overpasses.

For the given size box there are typically two Jason-2 passes (one ascending, one descending) and two Jason-3 passes, both of which are run thrice in a month, with about six AltiKa tracks through that box each month. However the temporal sampling is far from uniform: there is a 4-day period at the beginning (Days 6.5-10.5) and a 5-day period at the end with no observations, even with 3 altimeters. Using larger grid boxes will improve temporal sampling, but not allow the resolving of spatial changes in the wave climatology. The temporal sampling will be even worse in grid boxes containing significant land, as there will be fewer valid tracks. To ascertain the error associated with the gridding process, a mean Hs field for January 2017 was calculated separately for both Jason-2 and Jason-3 (see Fig. 52).

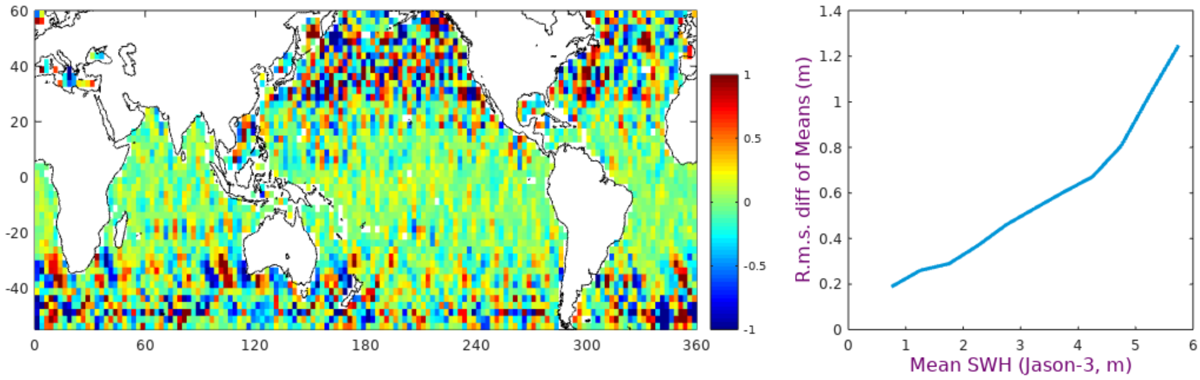

Figure 52 : a) Left panel shows difference in mean wave fields for January 2017 using either just Jason-2 data or just Jason-3. b) Right panel shows the r.m.s. difference of the Jason-2 and Jason-3 mean fields as a function of mean conditions. (Data used for this are from Dec. 2016 - Feb. 2017).

The difference in the gridded fields (Fig. 52a) matches the pattern of the mean Hs conditions (not shown), with the r.m.s. difference rising monotonically with wave height. This gridding error would be reduced by compiling data over larger grid boxes. Considering Jason-2 and Jason-3 to have near-identical performance, the uncertainty in their individual climatologies would be characterised by the curve in Fig. 52b divided by √2. Provided all altimeter estimates of Hs are harmonised (which is one of the aims of the Sea State CCI) then the gridding error associated with limited temporal sampling can be reduced by using multiple altimetric datasets.

4.1.2 Sigma0

For the development of monthly climatologies, it would be meaningless to average σ^0 values and then apply a wind speed algorithm. Instead the wind speed algorithm should be applied to the 1 Hz data and then averaged and gridded to provide a monthly product (L4).
Observations along a short segment of altimeter track will be synoptic and thus principally representing that particular wind environment. Typically further altimeter tracks through that box (see Fig. 51) are likely to sample a different section of the monthly wind conditions, and thus the variability of the mean meteorological conditions associated with that location and period. It is highly likely that no altimeter overpass will coincide with the most extreme conditions; however, assuming a log-normal distribution, the number of independent passes, gives an indication of the likely sampling error in the estimate of the mean conditions.

4.2 Estimation of Extremes

The characterization of the most extreme conditions (largest wave heights) in a region is of great societal relevance. The spasmodic sampling shown in Fig. 51b implies that major storms may pass unrecorded if they occur at inconvenient times. There is little that can be done concerning individual storms that were not measured; however by assuming that the nature of the long-term statistics is known (e.g. conformance to a Weibull distribution) then inferences can be made on the likely extremes. This will still not cover extraordinary events. Furthermore, in developing a long-term climatology of observed extremes, care must be taken to account for the greater likelihood of observing an extreme event when there are more altimeters in operation. Such an issue could be avoided by only using “2-satellite” operation (i.e. one in TOPEX/Jason orbit and one in ERS/Envisat orbit), but that involves a decision to discard more than half the altimetric data collected. The sparse sampling can also lead to a strong bias in parameters like 10-year return period (Stopa, 2016). One possibility to get away from this problem is to use a numerical model after properly quantile-quantile correcting the model biases using satellite data

Finally, a remark must be made about the effect of rain. Although the methodology for deriving wave height from altimeter returns is robust to most environmental effects, inhomogeneous patterns of atmospheric attenuation may seriously disturb the waveform shape and lead to wildly varying estimates of wave height. Figure 53 shows that the locations of sharp changes in derived Hs align with known patterns of rain. This is perhaps a simple but salutary reminder of the need for adequate quality control in the data passed towards the gridding process.

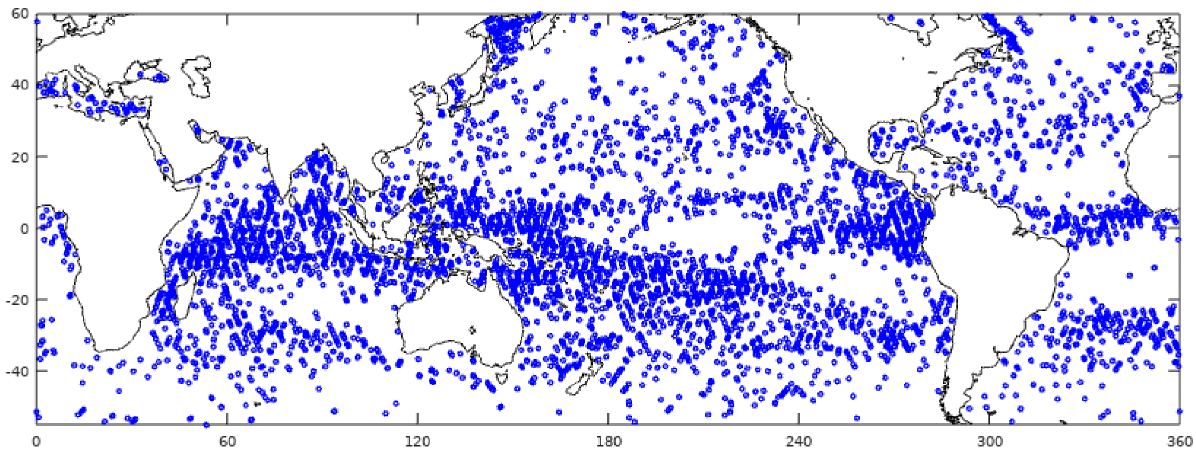

Figure 53 : Locations where σ_{Hs}>1.5m (see Fig. 3a for mean values). [Note this indicates occasional high values for Jason-3 during Dec. 2016-Feb. 2017, rather than the typical values there].

Although the pattern of the main rain bands is clearly apparent, these extreme values only represent 0.15% of the dataset. Careful rain-flagging is required as major storms with exceptional wave conditions occur preferentially in these regions of spurious Hs estimates due to the effect of rain.

5. Total Error

This document has detailed that there are a number of individual sources of error. An indicator of the total error can be gained by comparing altimetric data against some other trusted source. First we show a comparison with wave buoys, noting however that the square of the total mismatch error is the sum of the squares of the errors in the altimeter, and in the buoys and that due to their different spatiotemporal sampling of waves. In the subsequent sub-section, we explore use of the Triple Collocation Technique to partition the error between different sources.

5.1 Comparison of Altimeter Estimates with Buoys

Use of buoys for validation leads to an overestimate of the errors associated with altimeter records of wave height. This is because i) the buoy gives a point measure compared with an altimeter's areal average over the instrument footprint, ii) the buoy provides a temporal average (to compensate for its “point” nature) that may differ by half an hour from the time of the altimeter overpass, and iii) there are errors in the buoy measurement. The latter include the fact that buoys are not able to sample the whole wave spectrum, that calibrated buoys may have an error of order 2%, and that different buoy operators may have different procedures.. Finally, the buoy location may be sufficiently off track that it samples a different wave-field, especially as many buoys are in quite coastal locations. Usually buoy-altimeter comparisons are done far from the coast to avoid the effect of land on the altimeter signal and to minimise the likelihood of significant small-scale variations in the wave field. Thus the illustration in Fig. 54 is from the data gathered for the Round Robin, and shows the comparison for the 40 buoys showing the best agreement with the Jason-3 altimeter.

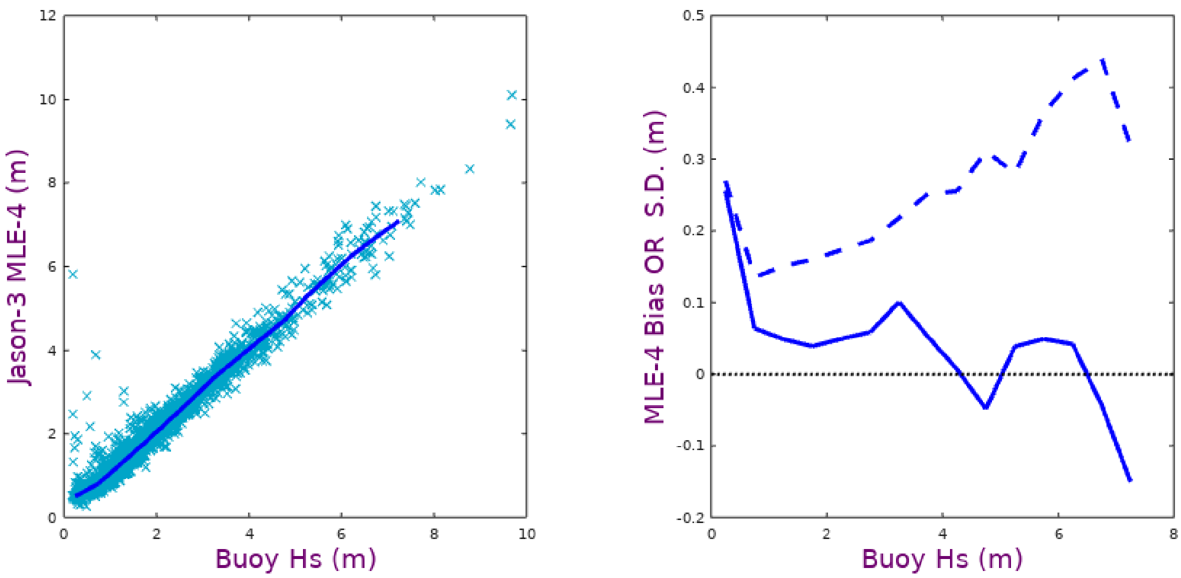

Figure 54 : a) Left panel shows scatterplot of matched up buoy and Jason-3 measurements. The buoy data have been smoothed over 3 hours and then linearly interpolated to the time of the overpass, whilst the altimeter values are the mean of the 51 20 Hz records nearest to the buoy. b) Right panel shows bias (solid line) and S.D. (dashed line) of MLE-4 as a function of wave height.

Despite having selected the buoys giving the best agreement, this is still an overestimate of the error in the altimeter value due to the reasons stated above. Figure 55 shows a comparison of S.D. as a function of SWH for the many algorithms evaluated in the Sea State CCI Round Robin exercise for altimetry.

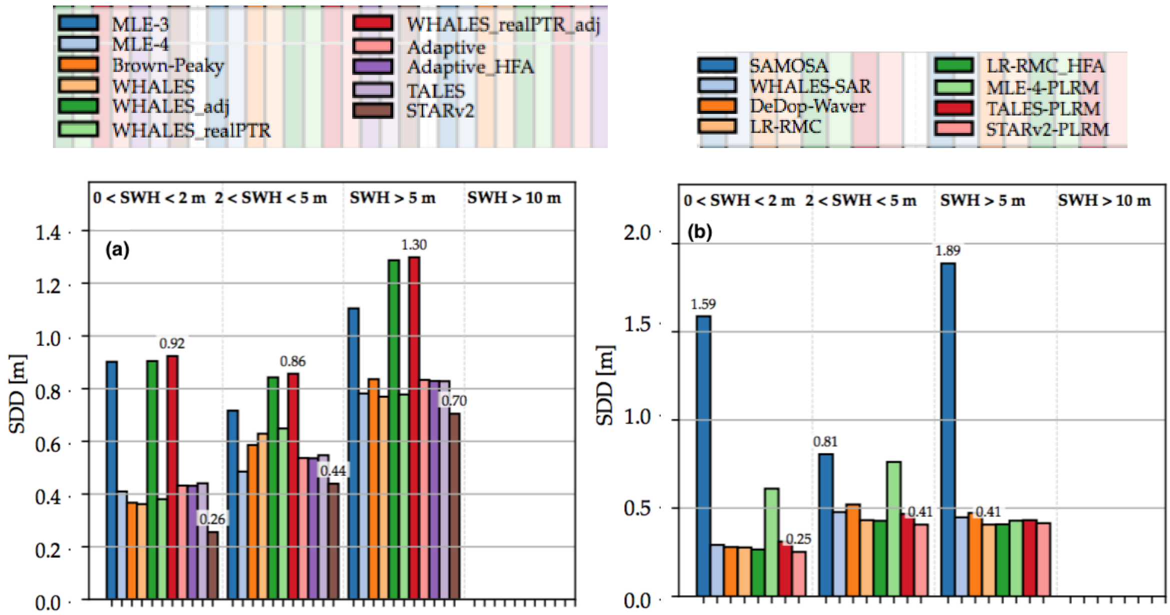

Figure 55 : Standard deviation of differences of various altimeter retrackers against in-situ buoy data as a function of SWH for a) Jason-3, b) Sentinel-3A respectively. *[Taken from Schlembach et al. (2020)].*

The RMSE values obtained from the comparison with buoys does depend very much upon the data flagging applied and the choice of buoys to be used.. If the 1 Hz average values are calculated from very few observations it is likely that these are all affected by land contamination of the waveforms. Requiring that there are a minimum of 5 or even 10 or 20 points for a mean to be defined removes those match ups beset with outliers and consequently improves both r^2 and RMSE (see Fig. 56a)

Further sub-selection of match ups according to whether buoy and altimeter track experience the same exposure to wave conditions removes pairing where one set of observations is in a sheltered location, and again leads to improved metrics (See Fig. 57a). Note, it is not simply a case that one type of comparison is wrong: open ocean comparisons with close match ups show the accuracy of the altimeter instrument (once buoy error is allowed for), whilst coastal results show the problems with contaminated waveforms, and the greater variability for poor match-ups highlights that altimeter values at the coast cannot be taken as representative of the conditions in a nearby coastal location with different exposure.

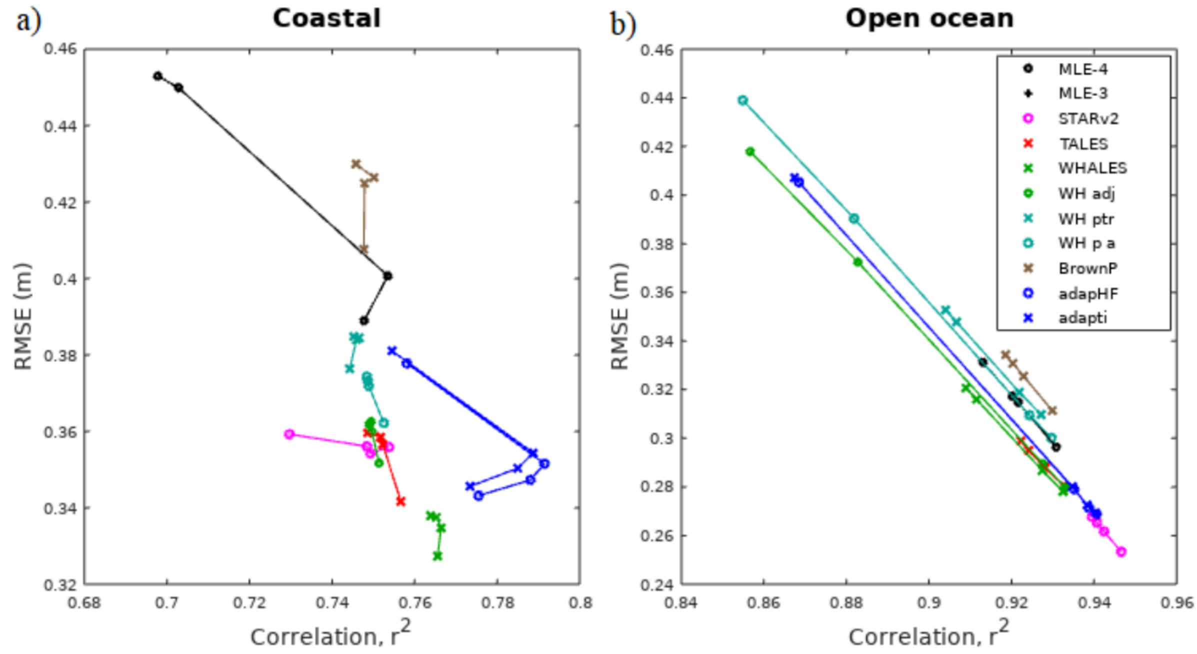

Figure 56 : Generally improved performance as minimum no. of valid measurements is increased from 1 (at top left) with low r^2 and high RMSR) to 5 to 10 to 20. Results are for various retrackers applied to Jason-3. [Taken from Quartly & Kurekin (2020)].

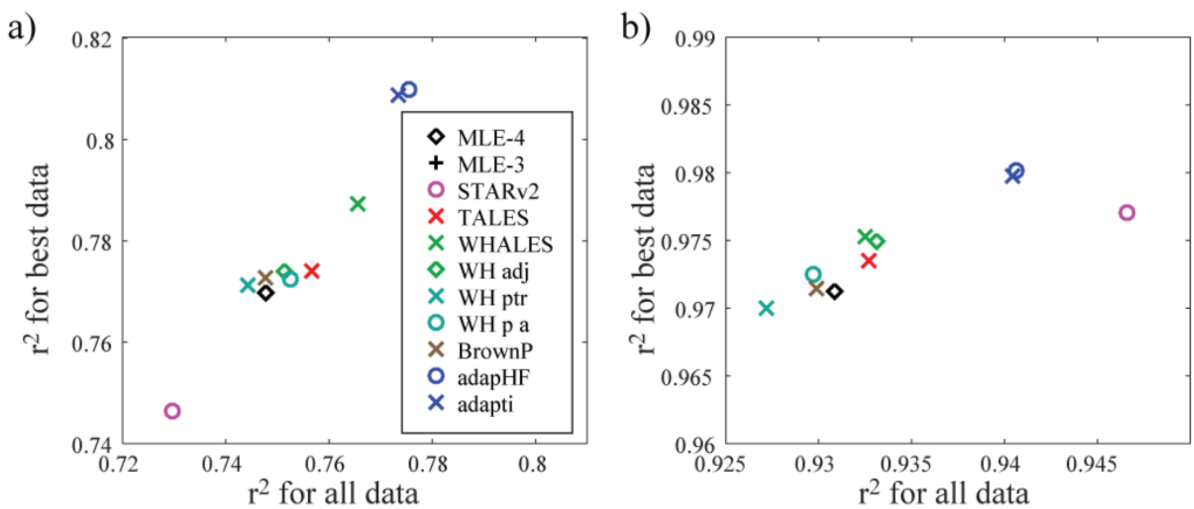

Figure 57 : Effect of match up selection on r^2 metric for (a) Coastal, (b) Open ocean. Results are for various retrackers applied to Jason-3; in all cases the r^2 value is higher (better match) when the comparison is restricted to the best aligned buoy-altimeter pairings.. [Taken from Quartly & Kurekin (2020)].

5.2 Comparison of SAR Estimates with Buoys

The SAR comparisons with buoys were conducted for 57 buoys found to be collocated with S1 WaveMode imagettes with a distance less than 50 km to imagette borders (see Fig. 58). This distance was chosen as optimal separation due to two conditions: the largest number of collocations with the acceptable local sea state variations. In fact, worldwide only 2 NDBC buoys are collocated closer than 2 km to S1 WV imagettes and only 15 buoys under 20 km.

The comparisons with buoys conducted in the scope of the round robin results in an RMSE of ~0.42 m for total SWH (see Fig. 59).

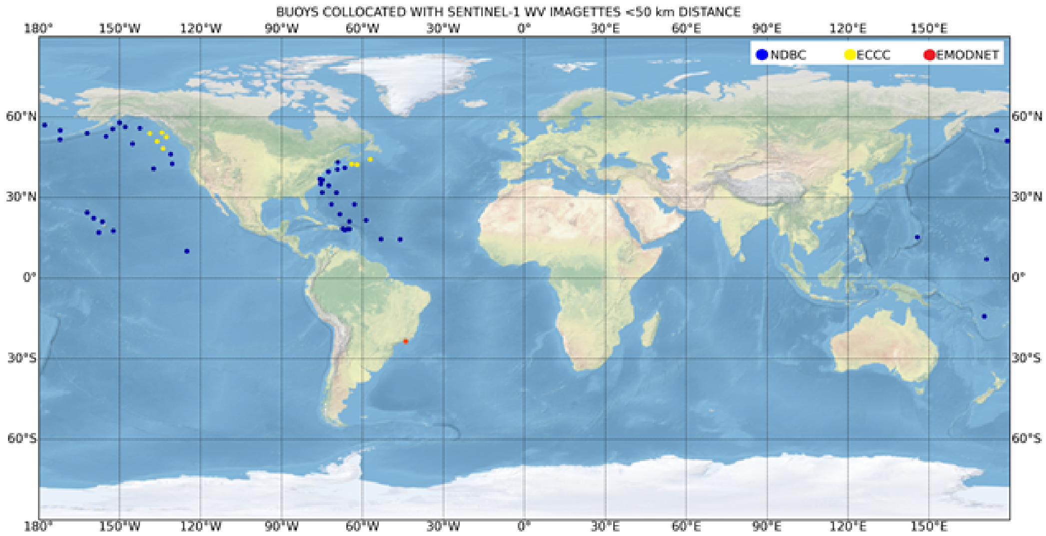

Figure 58 : 57 buoys collocated with S1 WV imagettes with distance less than 50 km to the imagette borders. There are 48 NDBC (blue), 8 ECCC (yellow) and 1 EMODNET (red) buoys.

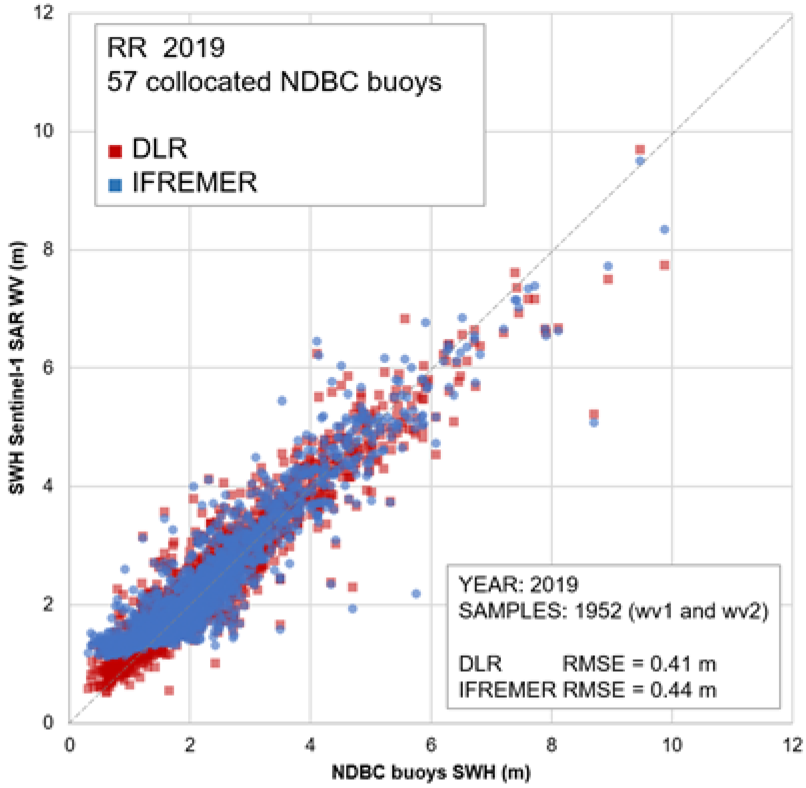

Figure 59 : Scatter plot for SWH estimates from SAR compared with NDBC measurements conducted for 2019. Averaged RMSE~0.42 m for total SWH.

5.3 Triple Collocation Technique: Comparison with Buoys and Models

The Triple Collocation Technique (TCT) is a statistical approach relying on the matchup of three totally independent estimates of a true field, and requiring that overall each of the estimates is unbiased with respect to each other, as it does not provide a means of estimating the bias of individual sources. It has been used to study sea surface height (Tokmakian and Challenor, 1999), sea surface temperature (O'Carroll et al., 2008) and metocean parameters (Caires and Sterl, 2003) amongst others. A key parameter choice is the amount of separation between buoy and altimeter measurements that is deemed acceptable. Abdalla and de Chiara (2017) showed, for wind speed, that results were robust if there were a "few thousand" match-ups. Abdalla et al. (2011) showed that this number could be achieved on an annual basis using a 50 km criterion with the available in situ network, but recommended relaxing the match-up threshold to 200 km, thus quadrupling the number of comparisons used. Their analysis (reproduced in Fig. 60) showed that for a separation of less than 50 km, averaging across all three altimeters investigated, the error in the satellite measurement would be ~0.16 m, rising to ~0.19 m if 200 km distance between altimeter and buoy is tolerated. Extrapolating the altimeter error line to zero distance gives an altimeter error of about 0.15 m (about 0.12 m for Jason-2 and Envisat and about 0.19 m for Jason-1). [Note the figure suggests that values about 0.02 m less than this would be pertinent if just Jason-2 and Envisat were considered; this *may* also be true for more recent altimeters such as Jason-3 and Sentinel-3A.]

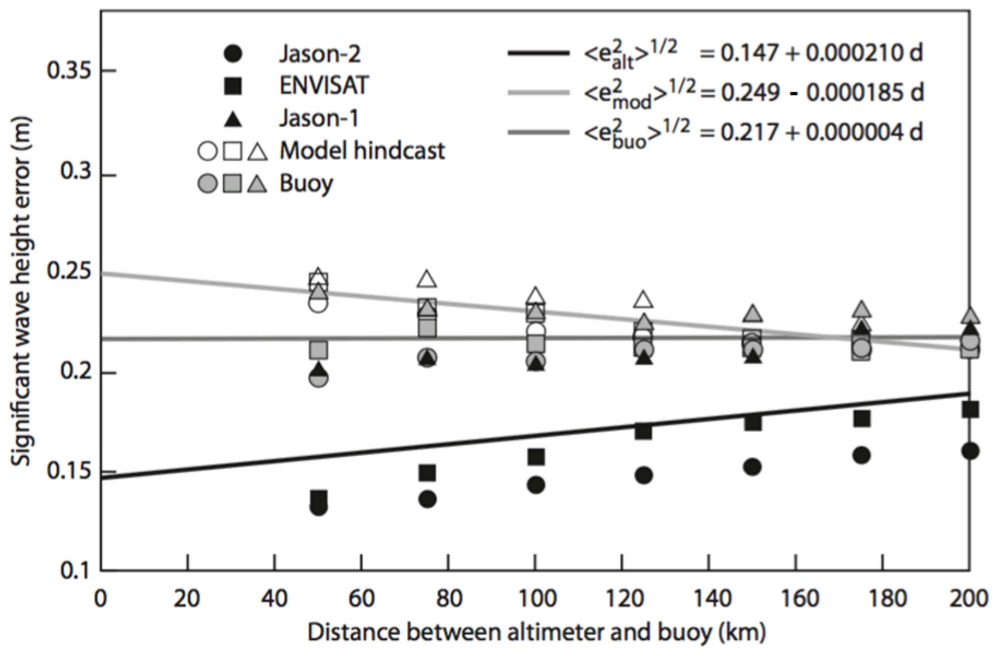

Figure 60 : The partition of errors from a Triple Collocation analysis, indicating that slightly greater errors are attributable to the altimeter as its separation from the buoys increases. [Taken from Abdalla et al. (2011)].

The direct buoy-altimeter comparison and the triple collocation technique offer slightly different perspectives on the error in the altimeter estimates. The former gives a mismatch error including buoy error, whereas the latter more clearly differentiates the errors due to various sources. However, TCT requires many more match-ups to give robust results and thus is less amenable to characterising the error in the coastal zone or as a function of wave height.

During phase 1 of the Sea State CCI project, Dodet et al (2022) applied the TC technique to SWH measurements from the 10 altimeters of the version1 dataset in combination with measurements from 122 in situ platforms and simulation results from the Ifremer global wave hindcast. With this extensive TC dataset, they were able to characterize the altimeter random errors at a yearly timescale and to compare these errors between the different missions (Fig. 61). They found that ERS-1 and ERS-2 showed the largest normalized errors (9.29% and 8.90%, respectively), while the Ka-band AltiKa instrument on-board SARAL showed the lowest normalized error (5.75%) and also the lowest interannual variability. The second lowest mean normalized error (6.69%) and standard deviation (0.26%) was obtained for the Envisat mission. Comparing the error between the different measurement systems, Hs measurements from Envisat, Jason-1, and SARAL present similar error levels than in situ observations, illustrating the excellent behavior of the radar systems on board these missions.

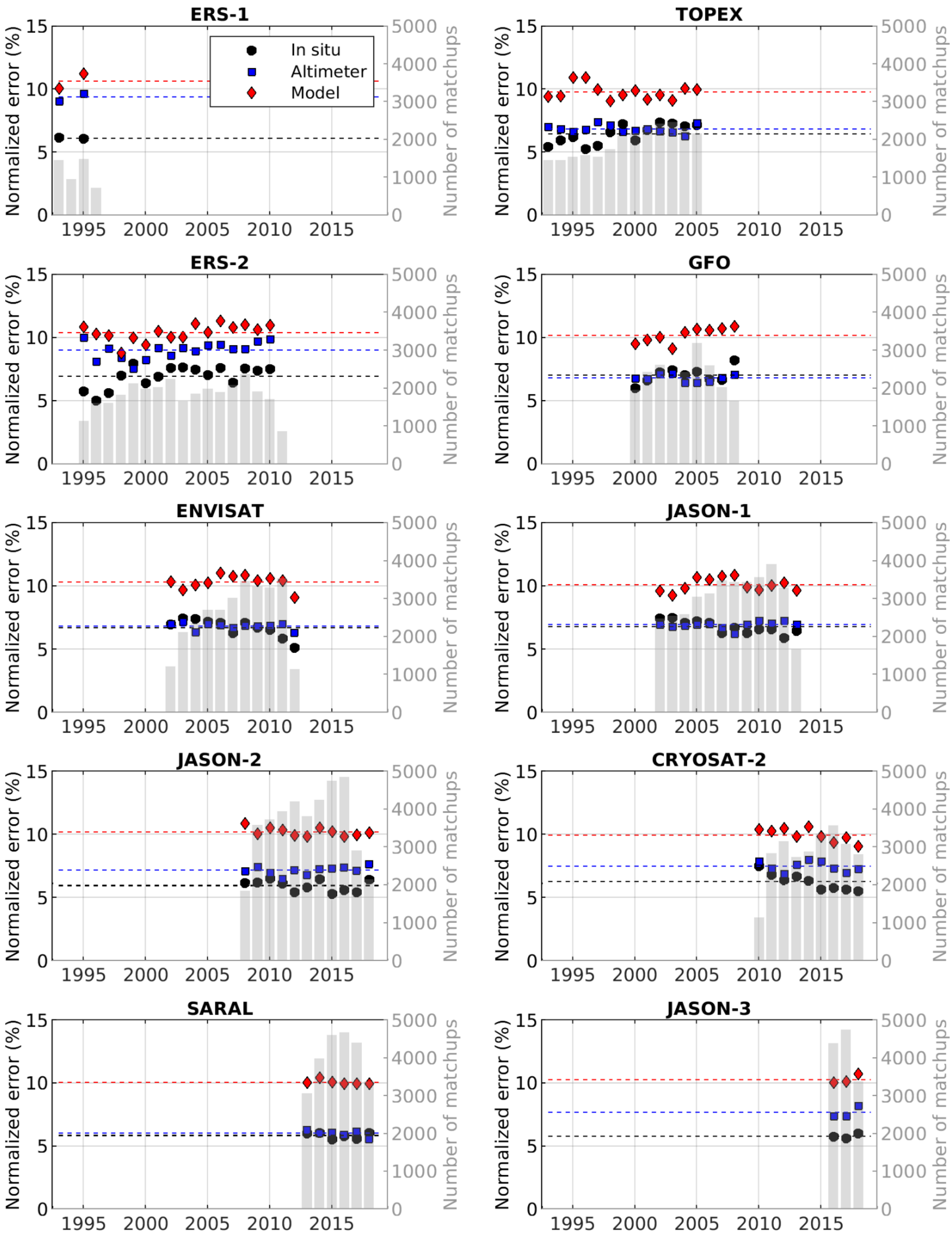

Figure 61 : Normalized error for in situ (black circles), altimeter (blue squares), and model (red diamonds) data, computed at yearly interval for each altimeter mission of the Sea CCI dataset. Gray bars indicate the number of matchups used to compute the error from the triple collocation technique. Dashed coloured lines represent the mean error values of each measurement system (from Dodet et al., 2022).

The TC technique also provided a means to investigate the behavior of the random errors of the three measurement systems as a function of sea states (Fig 62). The absolute random error in each measurement system could be modeled by a linear relationship over the ranges 0.5–8, 1.5–8, and 0–8 m, for the in situ, altimeter, and model systems, respectively. However, for Hs below 1 m the altimeter errors were found to be the highest and in situ errors to be the lowest. This increased error of altimeter measurement at low sea state can be directly linked to the limited range resolution of the instruments (~0.5 m for Ku-band radar), which is insufficient to adequately sample the steep leading edge of radar waveforms backscattered from calm seas, resulting in a noisy estimate of low Hs values.

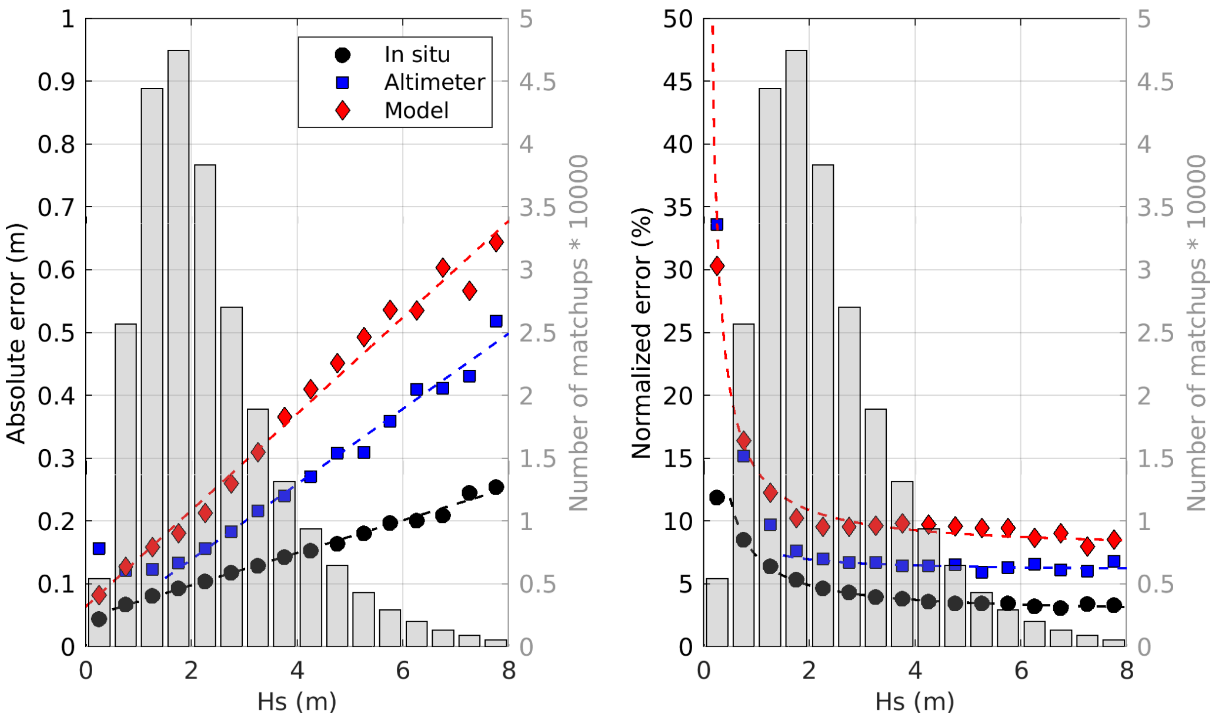

Figure 62 . (left) Absolute and (right) normalized errors for in situ (black circles), altimeter (blue squares), and model (red diamonds) data as a function of Hs. Gray bars indicate the number of matchups used to compute the error from the triple collocation technique (from Dodet et al., 2022).

5.3 S1 neural network approach in v5 from LOPS

An algorithm for estimating the uncertainty of geophysical quantities was developed in parallel with the algorithm for estimating geophysical parameters. This algorithm is applied to radar data estimated by the L1B processor of the SARWAVE processor (see SARWAVE ATBDs). The uncertainty of a geophysical variable is treated as an output variable of the machine learning algorithm and is estimated at the same time as the variable itself. Preliminary analyses of the relevance of the uncertainty estimated using this method have been carried out. They have shown very low sensitivity to different sea state conditions and do not, as they stand, provide relevant information on the quality of the associated geophysical quantity. Further studies should make it possible to estimate which input parameters have had the greatest influence on the inference of this uncertainty. These more in-depth studies should make it possible to modify the machine learning algorithm in order to improve the estimation of uncertainty.

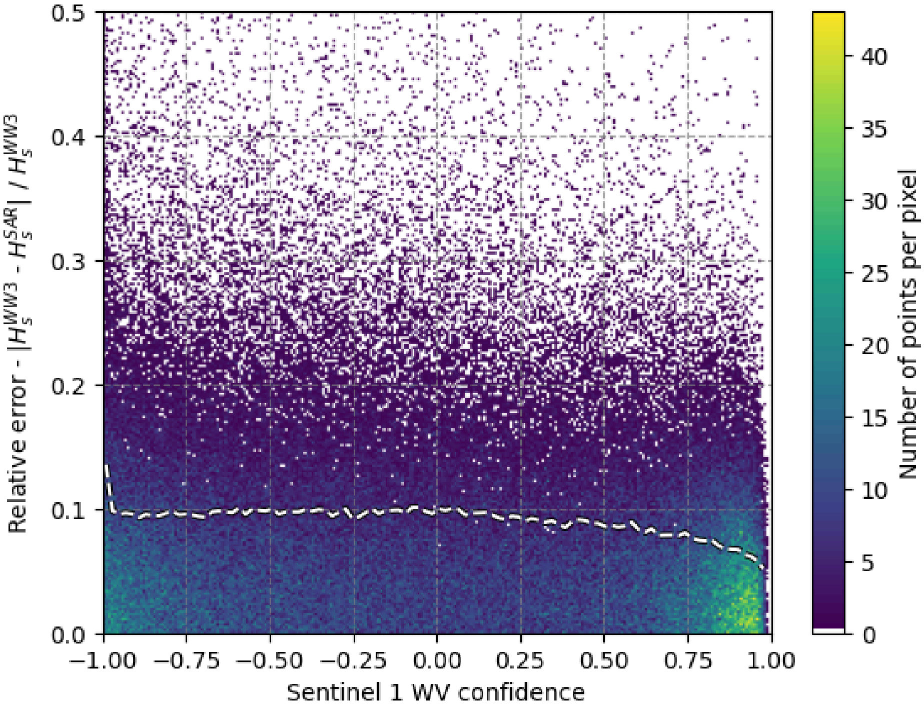

Figure 63 . Distribution of relative errors depending on the predicted confidence score. The flat appearance of the dotted line (average relative error) illustrates the little discriminatory power of the confidence score.

5.4 Insight from Climate Assessment Report

5.4.1 Data quality within the coastal zone

Given the considerable importance of sea states for coastal management and coastal risk mitigation, there is a large demand for high quality observations in the coastal zone. Since satellite observations (altimetry and SAR) can be contaminated by coastal features, it is relevant to investigate how close to the coast high quality satellite observations can be retrieved, and how measurement errors vary with the distance to the coast.

In order to examine these issues, calibrated data from Envisat and the Jason 1, 2 and 3 missions from the CCI L2P version 1.1 data product (Piollé et al. 2020) were evaluated. The results, presented in more details in the Climate Assessment Report (CAR), show that inshore waters (between 3 and 10 km) have higher SWH rms (noise) compared with the open ocean (between 43 and 50 km), the inshore waters and open ocean SWH rms noise tends to increase as the SWH increases, and the SWH shows a dramatic decrease in the number of observations less than 10 km from the coast (Fig. 63).

The results suggest that observations at a distance greater than 15 km from the coast do not suffer substantially from loss of data quality, as indicated by occurrence of rejection flags. This further implies that analysis of extreme SWH within 15 km is likely to be problematic, noting that the space-time sampling of extreme SWH from satellites is already fairly low. Without high quality in-situ observations, such as those provided by data buoys, to robustly validate satellite observations, or more in-depth understanding of the satellite rejection flags closer to the coast, **it may not be advisable to use the CCI L2P product V1.1 closer than 15 km to the coast**.

5.4.2 Long-term statistics

One of the key objectives of the Sea State CCI project is to robustly assemble multi-mission altimeter data in order to investigate sea state decadal variability (including trends) at global scale. Previous analysis of existing multi-mission altimeter products revealed significant trends of extreme significant wave height (trends of Hs90 up to 1 cm/year) over the period 1985-2018 (Young and Ribal, 2019). However, several factors (calibration method, reference dataset, sampling) may impact the accuracy of the computed trends. As a result, different sea state products may provide different trends.

As part of the Sea State CCI Product Assessment, the version 1 of the Sea State CCI dataset was compared with three other high-quality global datasets (Ribal and Young, 2019, hereafter, RY2019; ERA5; and CY46R1) using a consistent methodology (see Section 2.3 of CAR and Timmermans et al., 2020). These comparisons revealed similar spatial distributions but significant differences in the amplitude of the trends between the considered model and altimeter data. In particular the differences in the two altimeter-based products (RY2019 and CCI2019) were partly attributed to the different reference in-situ dataset and methodology used to calibrate the altimeter records.

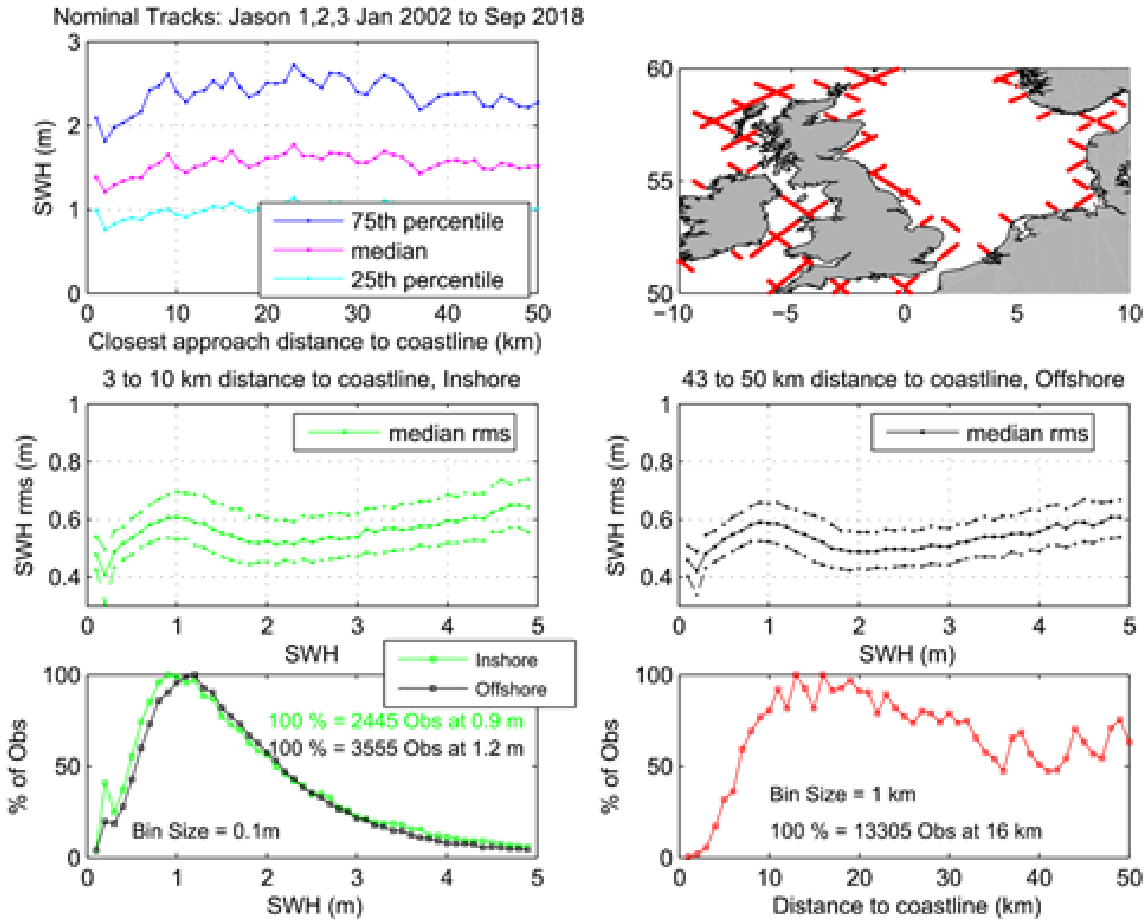

Figure 63 : Results from the CCI Sea State L2P product v1.1.using Jason 1, 2 and 3 for the United Kingdom and North Sea region. The SWH as a function of distance to the coast (top left panel), the location of the Jason tracks (top right panel). Inshore waters (3-10 km, middle left panel) and offshore water (43-50 km, middle right panel) showing SWH rms vs SWH. Lower left panel illustrates the percentage of good data as a function of SWH with respect to a bin size of 0.1m for both the inshore water (green) and offshore water (black). The lower right panel represents the percentage of good observations as a function of distance to the coast with respect to a bin size of 1 km.

Therefore, uncertainty on long-term trends related to the calibration method (e.g. due to changing reference in situ observations) should be further investigated and methods should be developed in order take it into account the total error budget of long-term statistics.

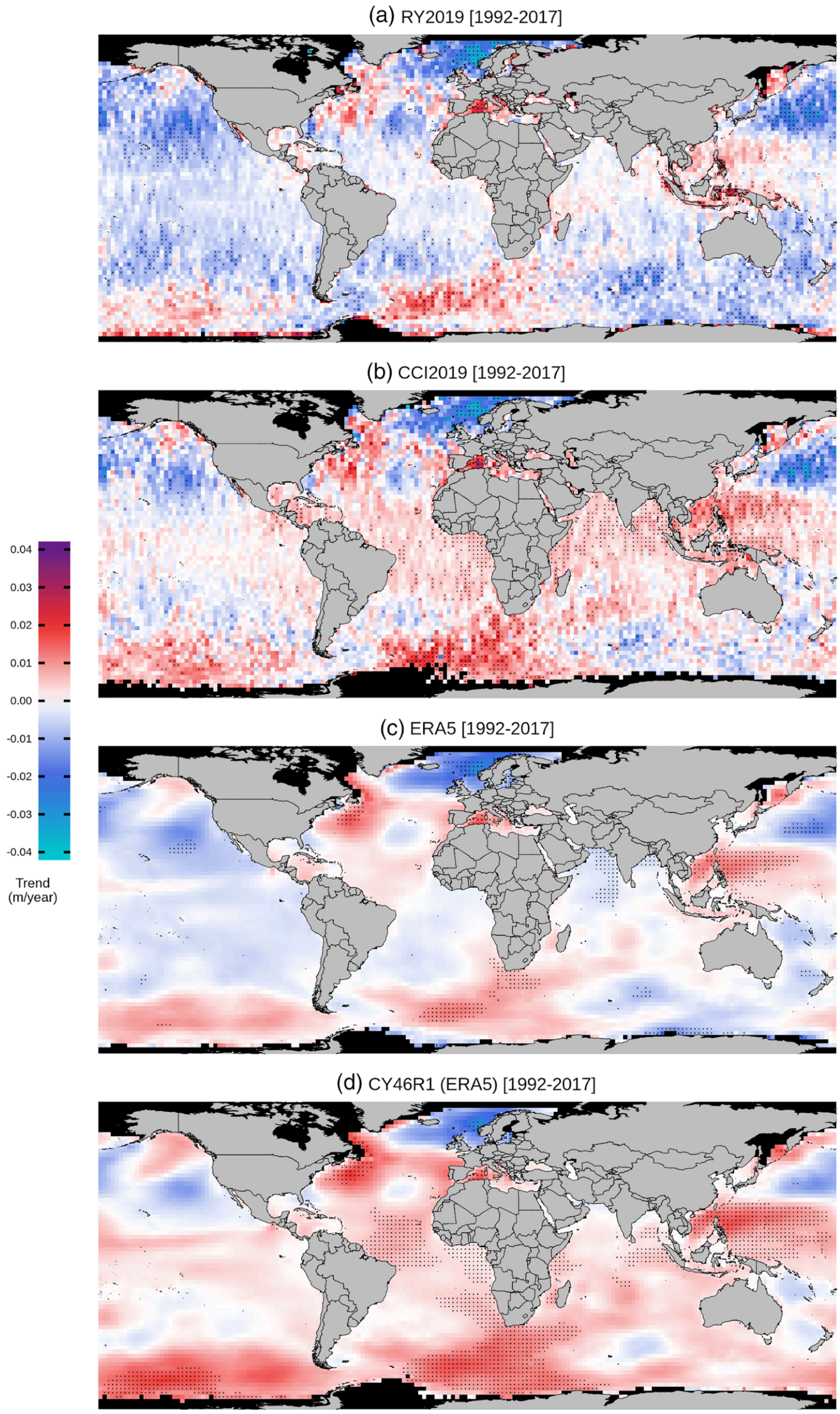

Figure 64 . Global distribution of JFM mean Hs trend estimates on a 2°× 2° grid over 1992–2017 for (a) RY2019, (b) CCI2019, (c) ERA5, and (d) CY46R1. Dots indicate grid cells where the trend coefficient is significant at the 95% level. [From Timmermans et al., 2020].

6. Summary

Radar measurements of wave conditions are little affected by other environmental parameters, except that heavy rain may affect the waveform shape and thus lead to occasional highly-localised errors. Estimates of Hs from SAR altimetry differ slightly from those from LRM processing, and so are not used in v4. They also show some sensitivity to relative direction of travel if the swell is of a very long wavelength (more than twice the along-track width of the delay-Doppler footprint).

There are errors associated with random instrumental noise, defects in algorithms (e.g. due to invalidity of some simplifying assumptions), and the limitations of the altimeter's spatio-temporal sampling of the true wave field. The table below brings together the values illustrated in the diagrams.

Effect	Magnitude	Notes
Fading noise on 20 Hz waveform	0.5-0.8 m	For Jason-3 MLE-4; can be improved with different retrackers
Fading noise on 1 Hz mean	0.12-0.20 m	For Jason-3 MLE-4. Up to 50% less for AltiKa. This can be improved with different algorithms (see Fig. 2)
Non-Gaussian PTR	±0.2 m	Only a pronounced error at low Hs (<1.0m, see Fig. 3)
SWH error from SAR (algorithm)	0.25 m ±0.01 m	S1 WV averaged for both algorithms (DLR and IFREMER), mean of wv1 and wv2, validated with CMEMS.
Partition SWH error from SAR (U-net)	0.63 m	Needs correction for a global bias; it is highest at high partition SWH (partition SWH > 6m)
Partition peak wavelength error from SAR (U-net)	30 m	Highest at low partition SWH (partition SWH < 1.5m)
Partition peak wave direction error from SAR (U-net)	27˚	Lowest (~16˚) at high partition SWH (partition SWH > 6m)
Sigma0 error due to AGC corrections	0.01 dB	Only noted as an issue for Jason-2
Along-track variability of sigma0	20 Hz: 0.07 dB 1 Hz: 0.016 dB	S.D. of estimates in 1 sec, larger for σ^0_{Ku} > 15 dB (see Fig. 45), but this is where wind speed is little affected. Standard error for 1Hz by dividing by √19; error likely to be much less when averaging over 50 km

Atmospheric correction to sigma0	0.02 dB	From consistency of MWR inversions and estimates from numerical weather models (see Fig. 48).
Inter-mission sigma0 biases	0.02 dB	Speculative estimate (to be assessed during Phase 2)
Wind speed error	0.1 m s^{-1}	Assuming total error in σ^0 in 0.03 dB and simple scaling between σ^0 and wind speed (see Fig. 50). Errors in wind speed model are **not** part of this assessment.
Error in gridded product (2.5˚ x 2.5˚ x 1 month)	0.15-0.80 m	For one altimeter, giving 6 independent transects during that time. Will reduce with more altimeter or larger boxes or periods. Error increases with mean Hs conditions (see Fig. 52b)
R.m.s. error for ~17 km section, compared with best buoys	0.20 m 0.35 m	This also includes the errors in the buoys and the difference in space and time sampling. Note mismatch error increases with Hs (see Fig. 54). Mismatch error in Round Robin, where almost all buoys are used (see Fig. 55) but significantly lower errors are achieved with selection of best pairings of buoys and tracks (Fig. 57).
Trends in L4 data	~ 0.01 m/yr	Fig. 60 shows derived regional trends from various differently derived datasets to differ locally by up to 0.02 m/yr. Incorporation or not of atmospheric modes leads to variations in derived long-term trend of ~0.005 m/yr (see Fig. 2.13 of CAR from Phase 1).

6.1 Recommended Error Specification in CCI v4 Products

A variety of products will be produced by the Sea State CCI, and users may wish to know the likely errors in each. This is not trivial, because (as highlighted in this document) the values vary with sea state, coastal proximity, technology, as well as retracker applied and any post-processing. Given the focus on the v4 product, the characteristics listed here are for WHALES or default (GDR) processing of LRM data with no subsequent filtering or smoothing applied. For simplicity, we characterise all as simple linear functions of Hs, with caveats noted.

High-rate (20 Hz) : This will be dominated by white noise, caused by fading. nse = 0.25m + 0.4Hs (LRM, Jason-3). There are slightly higher values at Hs<1m. This term can vary with instrument, but Jason-1, Jason-2, AltiKa and Envisat are likely to have similar values to Jason-3.

1 Hz data : Fading noise component still prominent (estimated magnitude is that o 20 Hz divided by √19 -- or √39, for AltiKa), but users may be looking to average the data themselves to **larger scales**, in which case the value they need is the altimeter error associated with buoy comparisons. This is estimated as 0.12m + 0.035 Hs (Fig. 16b).

Lastly there is a gridding error in producing a **L4 climatology**, which is purely associated with the spatio-temporal sampling. The analysis in section 4 was for 2.5˚ x 2.5˚ monthly boxes, which meant that a single altimeter would sample that domain about 6 times. On that basis, the representative error in a L4 product (based on Fig. 14b) is [0.15m + 0.12 Hs] * √(3/N), where N is the no. of independent transects. The actual proposed grid size for L4 products is 1x1 degree, so individual altimeters that are flying will not necessarily cover every box. Thus, there may be marked variations in the representativeness error for neighbouring cells, due to their very different sampling.

An assumption behind the above is that the data flagging is removing all the points affected by slicks, rain or land effects. Data in such situations may be of interest to users, but the associated errors will be larger and hard to quantify accurately. Such values may emanate from the WP on Validation.

For SAR (Sentinel-1 Wave Mode) the total noise of SWH estimation is 0.25 m. This is a mean value for both DLR and IFREMER algorithms, averaged over wv1 and wv2 validated with CMEMS. The value corresponds to the noise in ground truth data CMEMS (comparison CMEMS/WW3/NDBC results in RMSE around 0.25 m for buoy collocations). In detail, for wv1 the noise is 0.24 m and for wv2 noise is 0.26 m ± 0.01 m dependent upon algorithm and data. The SWH noise distribution can be approximated by 0.20 m + 0.04 * SWH· (see Fig. 65).

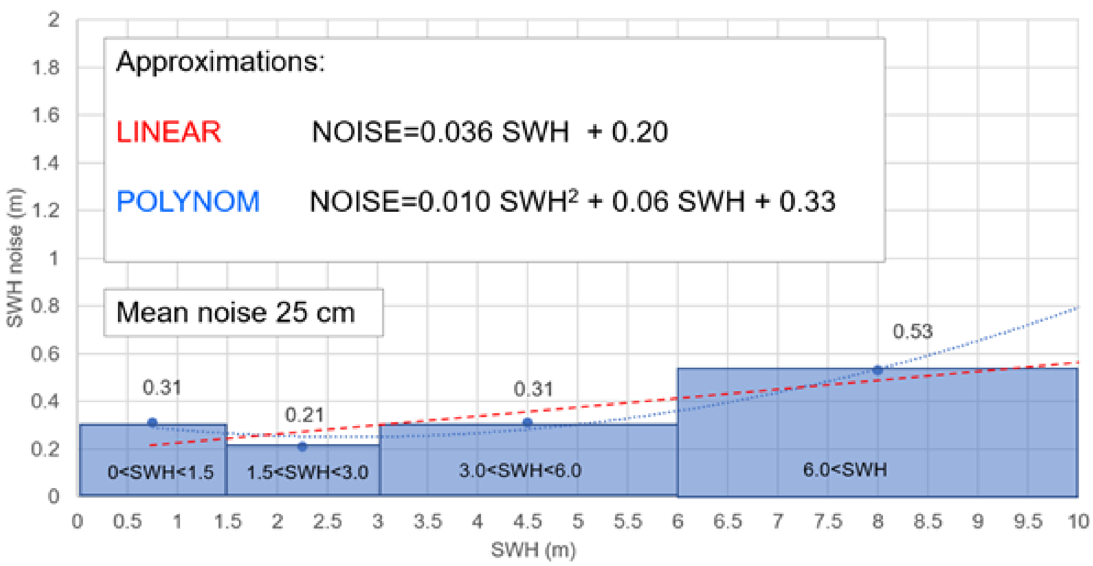

Figure 65 : SAR noise distribution with ground truth using CMEMS. Linear approximation NOISE=0.20+0.04 SWH.

7. References

Abdalla, S. 2012. Ku-band radar altimeter surface wind speed algorithm, Marine Geodesy, 35:sup1, 276-298. doi: 10.1080/01490419.2012.718676

Abdalla, S. and G. De Chiara, 2017. Estimating random errors of scatterometer, altimeter and model wind speed data, *J. Selected Topics in Applied Earth Obs. and Rem. Sens.*, 10:5, 2406-2414. doi: 10.1109/JSTARS.2017.2659220

Abdalla, S., P.A.E.M. Janssen, and J.-R. Bidlot, 2011. Altimeter near real time wind and wave products: Random error estimation, *Marine Geodesy*, 34:3-4, 393-406, doi: 10.1080/01490419.2011.585113

Caires, S., and A. Sterl, 2003, Validation of ocean wind and wave data using triple collocation, *J. Geophys. Res.*, 108 (C3), 3098, doi: 10.1029/2002JC001491

Camps-Valls, G., Tuia, D., Zhu, X. X., & Reichstein, M., 2021. Deep learning for the Earth Sciences: *A comprehensive approach to remote sensing, climate science, and geosciences* (p. 432). Wiley. doi: 10.1002/9781119646181

Collard, F., 2005: Algorithmes de vent et période moyenne des vagues JASON à base de réseaux de neurons. BO-021-CLS-0407-RF, Boost Technologies, 33 pp.

Dietterich, Thomas G., 2020. Ensemble methods in machine learning. International workshop on multiple classifier systems. Berlin, Heidelberg: Springer Berlin Heidelberg.

Dodet, G., S. Abdalla, M. Alday, M. Accensi, J. Bidlot, and F. Ardhuin, 2022. Error characterization of significant wave heights in multidecadal satellite altimeter product, model hindcast, and In situ measurements using the triple collocation technique. *J. Atmos. Oceanic Technol.* 39, 887–901. doi: 10.1175/JTECH-D-21-0179.1

Goan, E., and C. Fookes, 2020.Bayesian neural networks: An introduction and survey. Case Studies in Applied Bayesian Data Science: CIRM Jean-Morlet Chair, Fall 2018: 45-87.

Goodfellow, I, Y. Bengio, and A. Courville, 2016., *Deep Learning*. MIT Press

Gourrion, J., D. Vandemark, S. Bailey, B. Chapron, G. P. Gommenginger,, P. G. Challenor, and M. A. Srokosz, 2002. A Two-Parameter Wind Speed Algorithm for Ku-Band Altimeters. J. Atmos. Oceanic Technol., 19 (12), 2030-20048. doi: 10.1175/1520-0426(2002)019<2030:ATPWSA>2.0.CO;2

O'Carroll, A.G., J.R. Eyre, and R.W. Saunders, 2008. Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations, J. Atmos. Oceanic Technol., 25, 1197-1207. doi: 10.1175/2007JTECHO542.1

Piolle J.F., G. Dodet and E. Ash, 2020. Sea State CCI Product User Guide v 1.0

Pleskachevsky, A., B. Tings, S. Jacobsen, S. Wiehle, E. Schwarz, and D. Krause, 2024. A system for near-real-time monitoring of the sea state using SAR satellites, *IEEE Trans. Geosci. Rem. Sens.*. 62, 1-18, 2024, art no. 5219018, doi: 10.1109/TGRS.2024.3419582

Quach, B., Y. Glaser, J.E. Stopa, A.A. Mouche, P Sadowski, 2020. Deep learning for predicting significant wave height from synthetic aperture radar, *IEEE Trans. Geosci. Rem. Sens*. 59 (3), 1859-1867. doi: 10.1109/TGRS.2020.3003839

Quartly, G.D.,. 2009. Improving the intercalibration of σ^0 values for the Jason-1 and Jason-2 altimeters. *IEEE Geosci. Remote Sens. Lett*. 6(3):538–542. doi: 10.1109/LGRS.2009.2020921

Quartly, G.D., 2025. The intertwined factors affecting altimeter sigma0. Remote Sens. 2025, 17, 3776 (21pp.). doi:10.3390/rs17223776

Quartly, G.D., F. Nencioli, M. Raynal, P. Bonnefond, P.N. Garcia; A. Garcia-Mondéjar, A. Flores de la Cruz, J.-F. Cretaux, N. Taburet, M.-L. Frery, M. Cancet, A. Muir, D. Brockley, M. McMillan, S. Abdalla, S. Fleury, E. Cadier, Q. Gao, M.-J. Escorihuela, M. Roca, M. Bergé-Nguyen, O. Laurain, J. Bruniquel, P. Féménias, and B. Lucas, 2020. The roles of the S3MPC: Monitoring, validation and evolution of Sentinel-3 altimetry observations, *Remote Sens*. 2020, 12, 1763. doi: 10.3390/rs12111763

Quartly, G.D. and A.A. Kurekin, 2020. Sensitivity of altimeter wave height assessment to data selection. *Remote Sens*. 2020, 12, 2608. doi: 10.3390/rs12162608

Reichstein, M., Camps-Valls, G., Stevens, B. et al., 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566, 195–204. doi: 10.1038/s41586-019-0912-1

Ribal, A., Young, I.R., 2019. 33 years of globally calibrated wave height and wind speed data based on altimeter observations. *Scientific Data* 6, 77. doi: /10.1038/s41597-019-0083-9

Schlembach, F.; M. Passaro, G.D. Quartly, A. Kurekin, F. Nencioli, G.; Dodet, G. J.-F. Piollé, F. Ardhuin, J. Bidlot, C. Schwatke, F.; Seitz, P. Cipollini, and C. Donlon, 2020. Round robin assessment of radar altimeter Low Resolution Mode and Delay-Doppler retracking algorithms for significant wave height. *Remote Sens*. 2020, 12, 1254. doi: 10.3390/rs12081254

Shorten, C., and Taghi M. Khoshgoftaar, 2019. A survey on image data augmentation for deep learning. Journal of Big Data 6.1, 1-48.

Stopa, J., 2016. Wave climate observed from altimeters: Trends and extremes. *Poster at SAR workshop, OSTST 2016* (La Rochelle, 31 Oct-4 Nov 2016).

Timmermans, B.W., Gommenginger, C.P., Dodet, G., Bidlot, J.-R., 2020. Global wave height trends and variability from new multimission satellite altimeter products, reanalyses, and wave buoys. *Geophys. Res. Lett*. 47, e2019GL086880. doi: 10.1029/2019GL086880

Tokmakian, R., and P.G. Challenor, 1999. On the joint estimation of model and satellite sea surface height anomaly errors, *Ocean Modell*. 1,. 39– 52. doi: 10.1016/S1463-5003(99)00006-2

Young, I.R., and A. Ribal,, 2019.. Multiplatform evaluation of global trends in wind speed and wave height. *Science* 364, 548. doi: 10.1126/science.aav9527