



**sea state**  
cci

## End-to-end ECV Uncertainty Budget (E3UB)

version 3.0 8 Sept 2021

## Contents

<b>List of Acronyms</b>	<b>4</b>
<b>1. Introduction</b>	<b>5</b>
<b>2. Errors in the Instruments / Algorithms</b>	<b>6</b>
2.1 Altimeter (LRM)	6
2.2 Altimeter (DD)	9
2.3 Synthetic Aperture Radar	11
<b>3. Error Propagation</b>	<b>17</b>
<b>4. Uncertainty in L4 Gridded Products</b>	<b>18</b>
4.1 Uncertainty in the Mean	18
4.2 Estimation of Extremes	19
<b>5. Total Error</b>	<b>21</b>
5.1 Comparison with Buoys	21
5.2 Triple Collocation Technique: Comparison with Buoys and Models	25
5.3 Insight from Climate Assessment Report	26
<b>6. Summary</b>	<b>29</b>
6.1 Recommended Error Specification in CCI v2 Products	30
<b>References</b>	<b>32</b>

Author	Approved	Signature	Date
Graham Quartly and contributors	Ellis Ash		8 Sept 2021
<b>ESA Acceptance</b>			

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

## List of Acronyms

ADP	Algorithm Development Plan
ATBD	Algorithm Theoretical Basis Document
CAR	Climate Assessment Report
cci	Climate Change Initiative
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
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
PVASR	Product Validation and 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 three of the End-to-end ECV Uncertainty Budget (E3UB) for **Sea\_State\_cci**, deliverable 2.3 of the project.

The calculation of significant wave height ( $H_s$ ) 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 after 3 years of the *Sea State CCI* project.

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 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.

#### 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  $\sigma_{H_s}$  (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  $\sigma_{H_s}$  as a function of  $H_s$  for Jason-3.

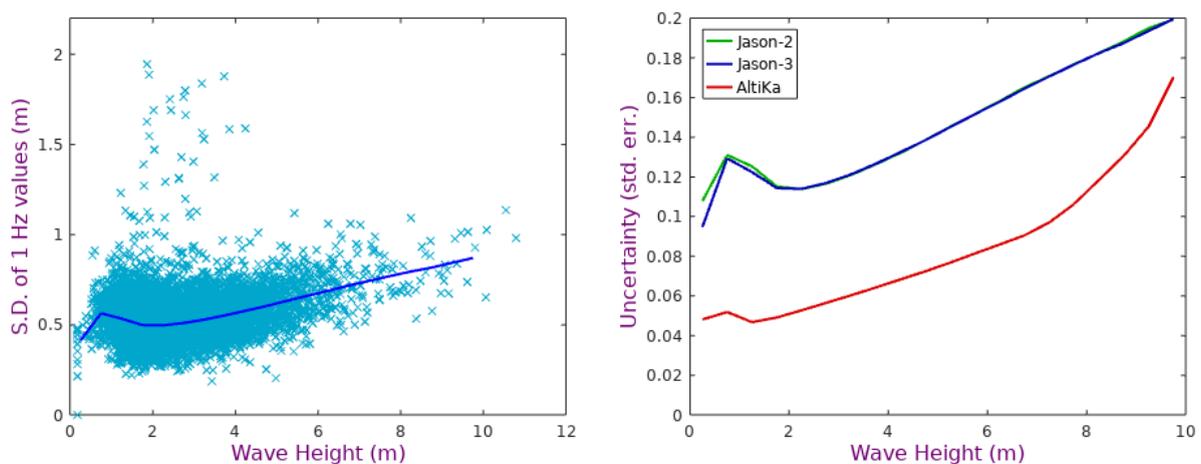
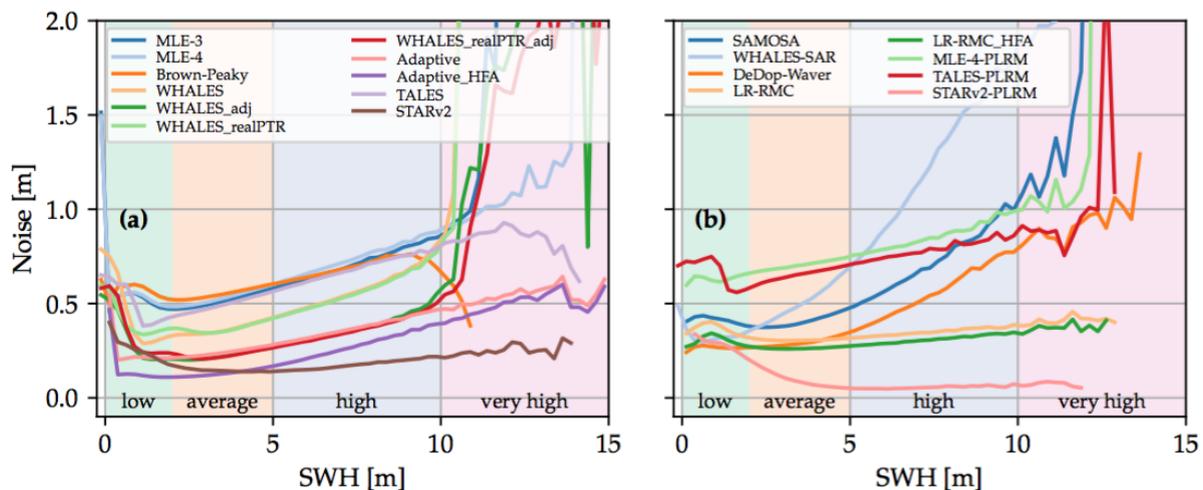


Figure 1 : a) Left panel shows a scatter plot of  $\sigma_{H_s}$  as a function of  $H_s$  for Jason-3's MLE-4-based estimates during Jan. 2017. The cyan crosses only show a small subset of the points, but the blue line shows the mean relationship derived from all data (averaged in 0.5 m bins). b) Right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

The analysis in Fig. 1a shows the results for the standard MLE-4 algorithm in the Jason-3 GDRs. The mean value of  $\sigma_{H_s}$  is  $\sim 0.5$  m at low wave heights and increases with  $H_s$ . 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  $\sigma_{H_s}$  for  $H_s < 0.3$  m 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 retrackerers 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 retrackerers 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  $\sigma_{Hs}$  as  $H_s$  approaches zero is less pronounced.



*Figure 2 : Noise level of the individual retrackerers 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)].*

### 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,  $\sigma_C$ :

$$\sigma_c^2 = \sigma_p^2 + (H_s/2c)^2 \quad (1)$$

where  $\sigma_p$  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  $\sigma_p$  (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  $H_s$  term will dominate  $\sigma_c$  for  $H_s > 2\text{m}$ , the concerns about the actual shape and width of the PTR are only pertinent for low  $H_s$  conditions. Figure 3a shows the inferred instrument correction on Jason-3 to compensate for the real PTR shape.

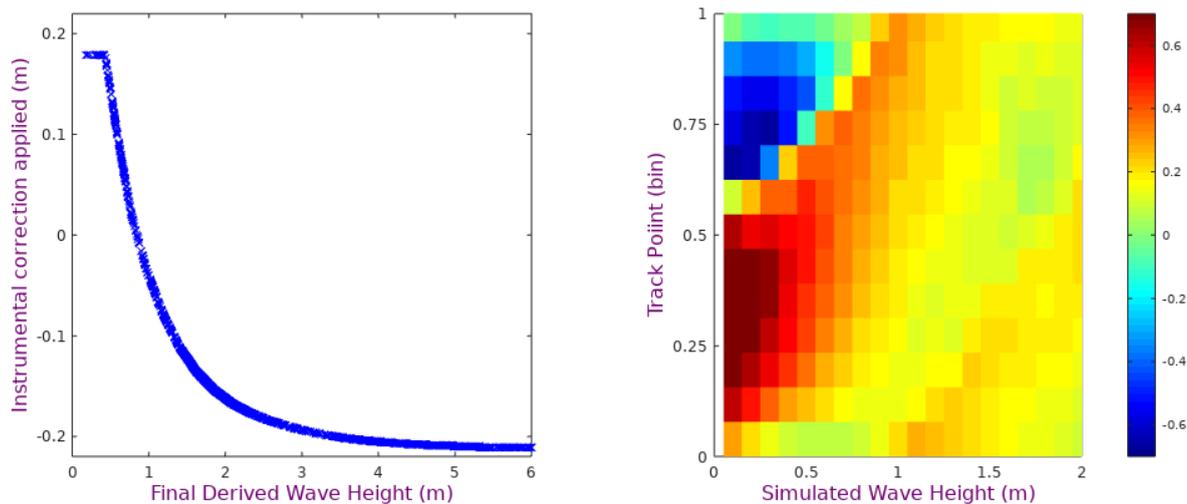


Figure 3 : a) Left panel shows LUT correction for MLE-4 algorithm applied to Jason-3. b) Right panel shows mean bias observed using simulation results for the WHALES algorithm, with the bias varying according to both wave height and the position of the leading edge.

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  $H_s$  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  $H_s$ , 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).

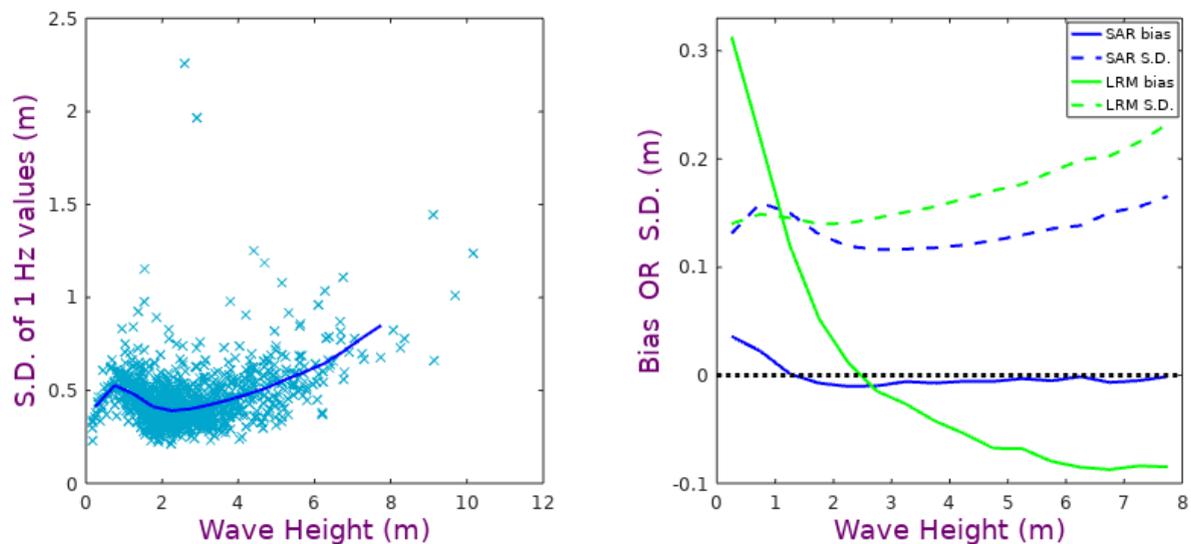


Figure 4 : a) Left panel shows  $\sigma_{H_s}$  (S.D. of values in a 1-second interval) as a function of mean  $H_s$ . Data are from Sentinel-3A cycle 033. Cyan crosses show a subset of the points, the blue line shows the mean relationship derived from all the data in the cycle. b) A comparison of S3B 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  $H_s$  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  $H_s$  to almost -0.1m at high  $H_s$ , 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).

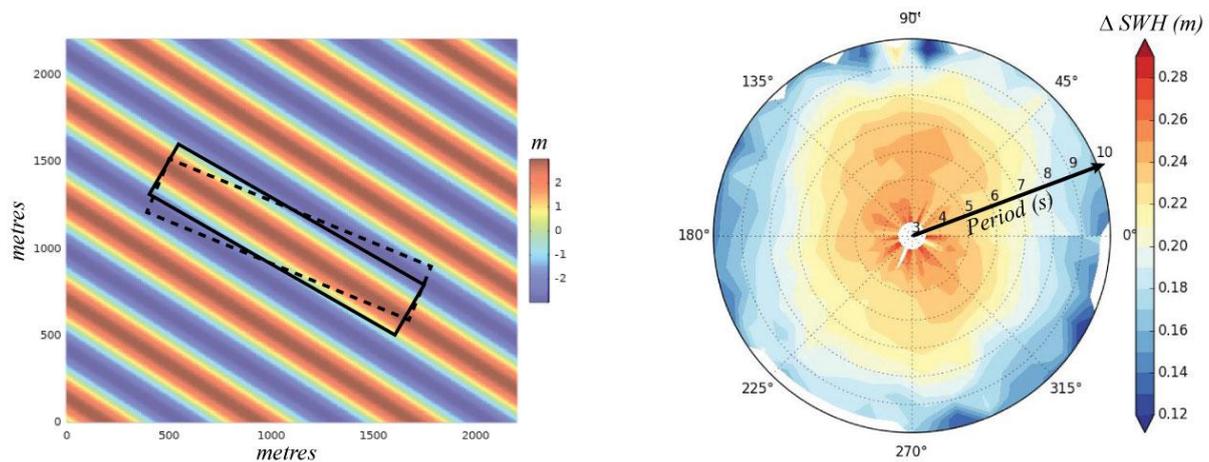


Figure 5 : a) Left panel shows schematic of a simple long wavelength swell field (colours indicating the height of the water surface), with the narrow SAR mode footprint almost aligned with direction of swell. b) Difference between SAR and PLRM estimates of wave height as a function of angle between 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 ( $0^\circ$  and  $180^\circ$  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  $H_s$  value; however this does not imply that the bias for 1 Hz estimates is simply a linear function of  $H_s$ , 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.

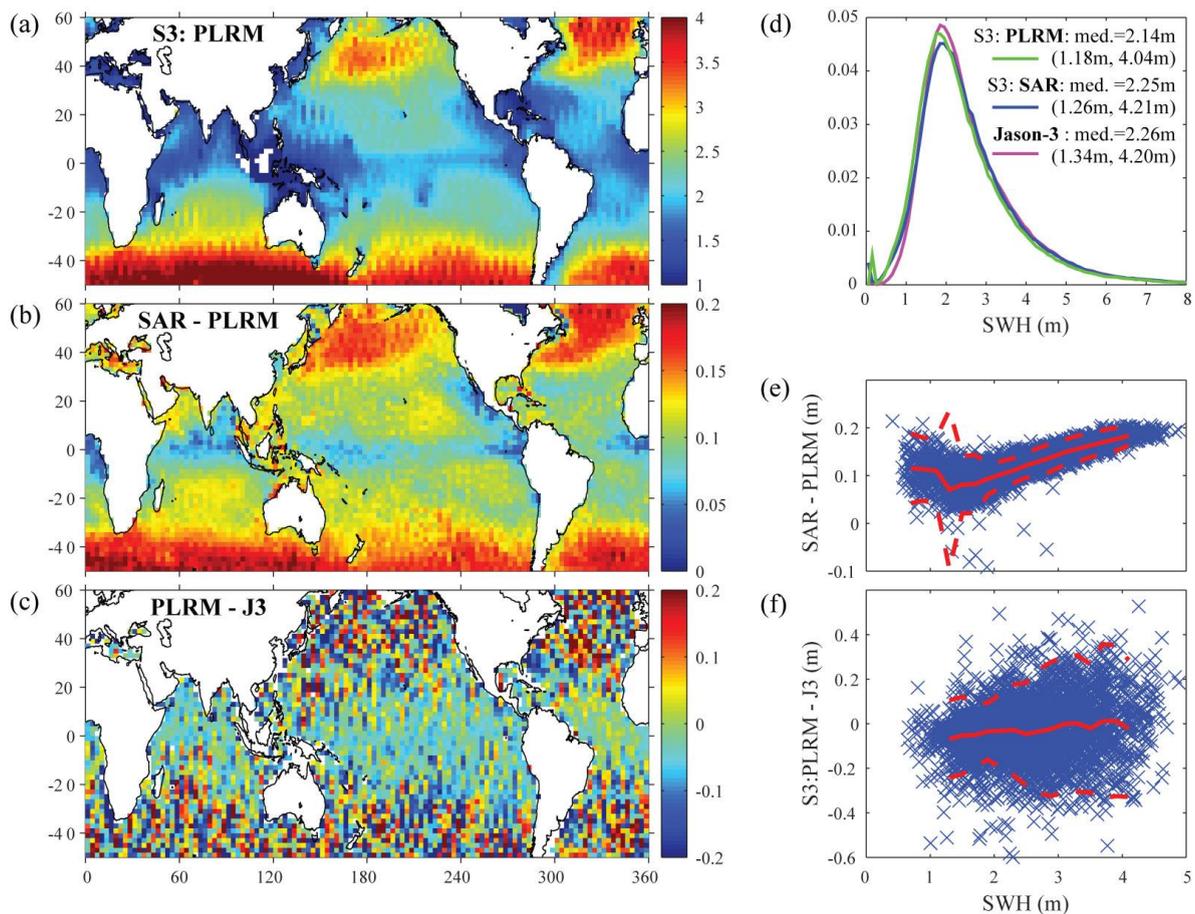


Figure 6 : Comparison of altimeter SWH data for 2019. a) Average of all 2019 for Sentinel-3 PLRM data b) Mean bias of Sentinel-3 SAR mode data to those from PLRM c) Difference of mean fields from S3:PLRM and Jason-3 (LRM) d) Probability distribution functions of SWH data for Jan. 2019. Numerical values 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  $\pm 2$  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)].

### 2.3 Synthetic Aperture Radar

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  $T_{m1}$ ,  $T_{m2}$ , 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 DLR method, around 1% of data are discarded (flagged) by atmospheric and other artefacts. For example, under strong winds, the sea surface can be completely destroyed 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  $\sim 0.25$  m for general conditions which, however, can reach an RMSE of  $\sim 0.5$  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 less 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).

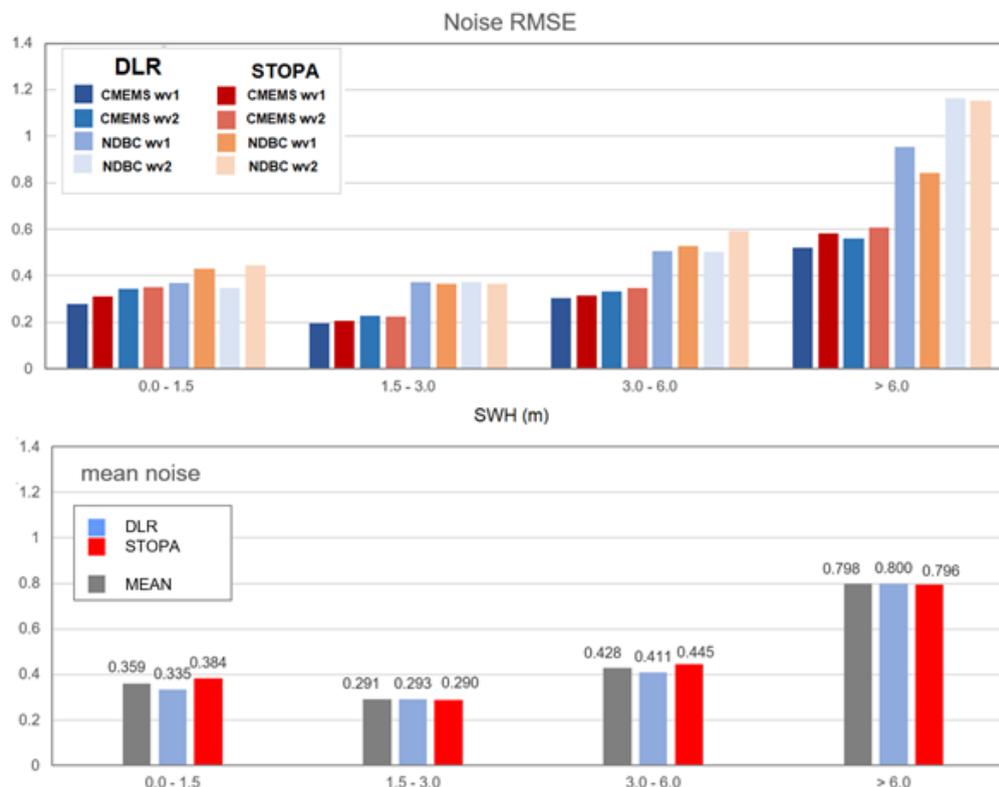


Figure 7 : RMSE distribution for four sea state domains and mean RMSE noise estimated. The mean values in the second line were built with weighting of 0.5 for both CMEMS and NDBC.

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  $-60^{\circ} < \text{LAT} < 60^{\circ}$  in order to avoid ice coverage. However, as analyses showed, sea ice can be encountered until  $-60^{\circ} < \text{LAT} < 55^{\circ}$  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. In case the validation area is reduced to  $55^{\circ} < \text{LAT} < 60^{\circ}$ , the resulting total RMSE improves by ~1.5 cm.

Fig. 8 shows the distribution of squared difference between SWH estimated from 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  $\text{LAT} < -58^{\circ}$ . This inconsistency is due to sea ice.

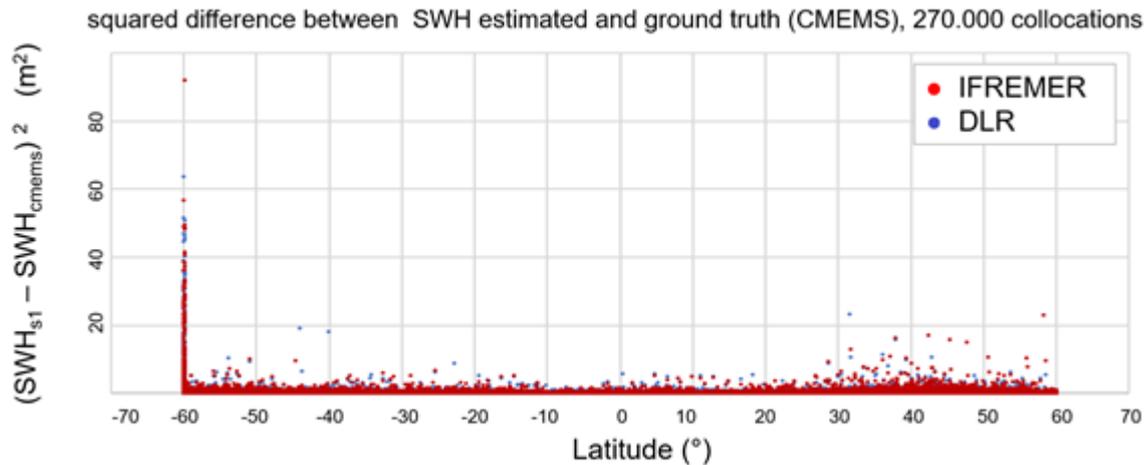


Figure 8 : Distribution of squared difference between SWH estimated from S1 and ground truth (CMEMS). For both algorithms, the ice coverage impact is quite visible for  $LAT < -58^\circ$ . The complete comparison was done for  $-60^\circ < LAT < 60^\circ$  in order to avoid ice coverage. As can be seen 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 a 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.

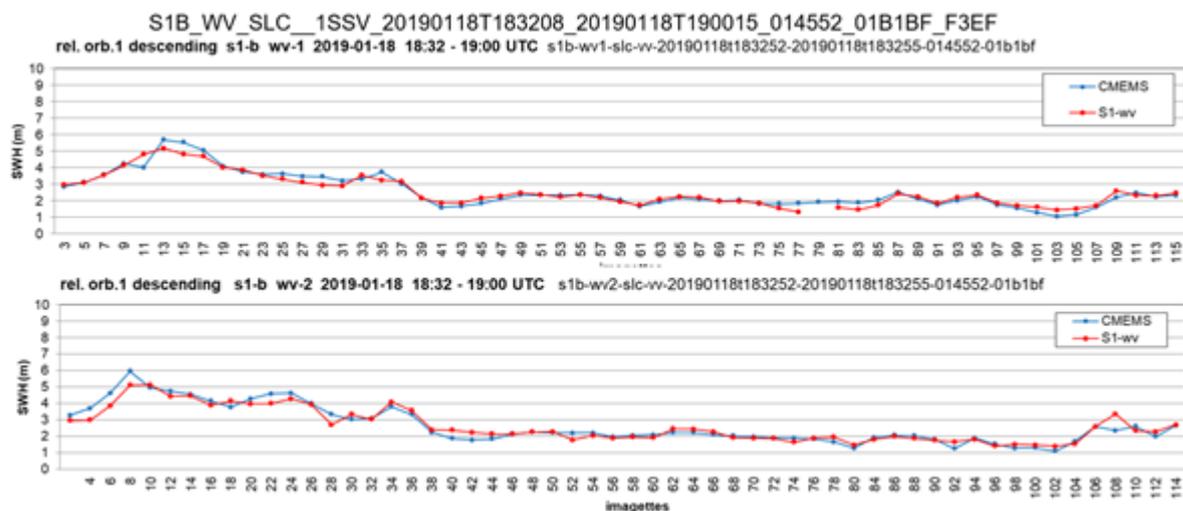


Figure 9 : An example of an along-orbit comparison for one long overflight of around 12.000 km (product ID S1B\_WV\_SLC\_1SSV\_20190118T183208\_20190118T190015\_014552\_01B1BF\_F3EF) plotted separately for wv1 (odd number imagettes on the first graph, 200 km between each points) and wv2 (even number imagettes on the second graph) vignettes. Red points mean DLR SWH and blue are CMEMS SWH temporally interpolated 3h model outputs.

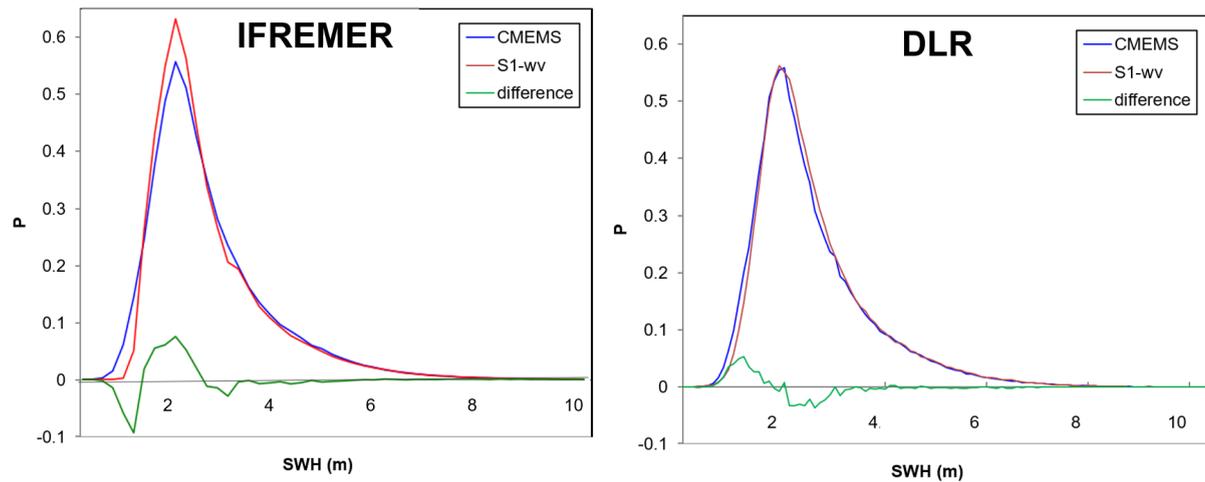


Figure 10 : P.d.f. of SWH distribution for CMEMS and estimated from S1 WV imagettes. a) Ifremer (Stopa) algorithm, b) DLR.

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) built for 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.

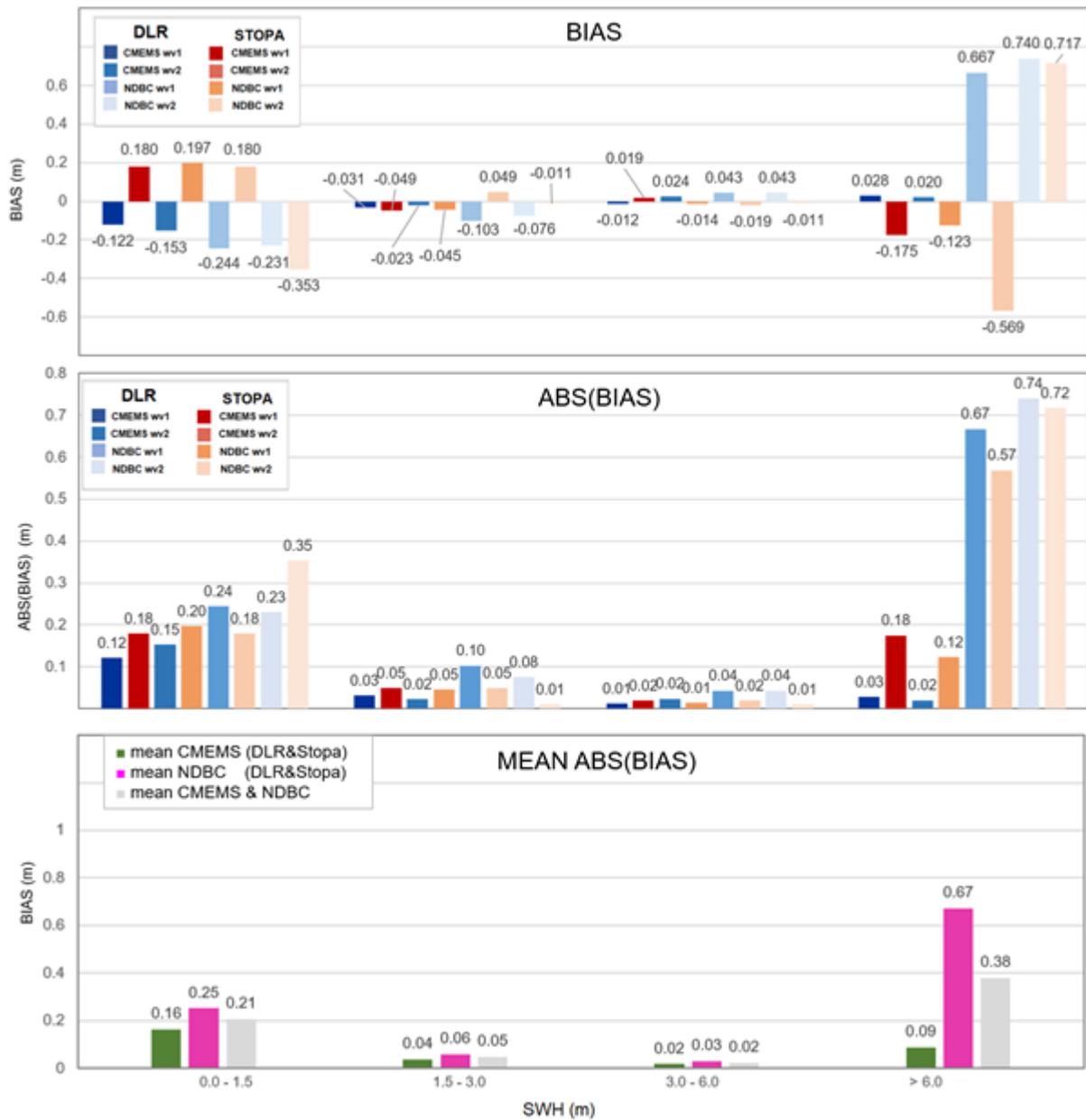


Figure 11 : Bias of both DLR and Ifremer (Stopa) algorithms (top graph), ABS(BIAS) (centre graph) and mean for CMEMS and NDBC for each sea state domain (bottom graph). Grey bars show averaged ABS(BIAS) for CMEMS and NDBC with weighting of 0.5.

### 3. Error Propagation

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. 12, 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.

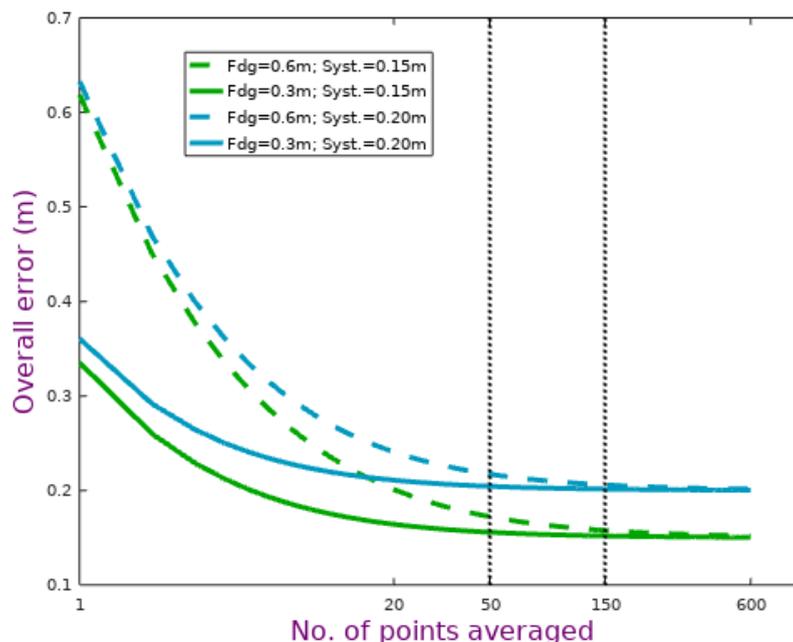


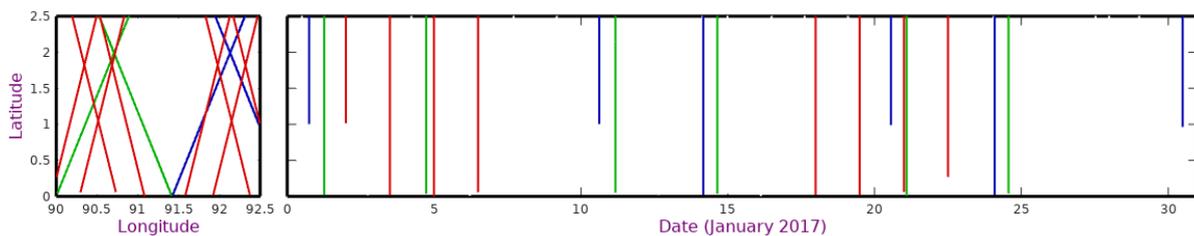
Figure 12 : Illustration to show that error in altimeter estimates due to fading noise dominates for scales of 1 Hz (20 obs.) and finer, but systematic errors are more important at larger scales. Dashed line at 50 pts represents the averaging scale used in buoy comparisons, and that at 150 pts equates 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.

## 4. Uncertainty in L4 Gridded Products

### 4.1 Uncertainty in the Mean

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  $2.5^\circ \times 2.5^\circ$  boxes (as this is close to the longitudinal spacing of Jason tracks) for a monthly period (see Fig. 13).



*Figure 13 : 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. 14).

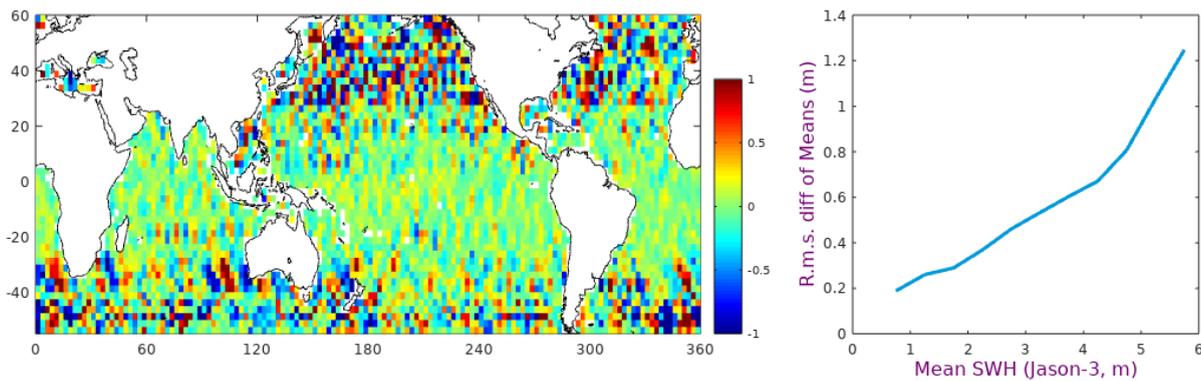


Figure 14 : a) Left panel shows difference in mean wave fields for January 2017 using either just Jason-2 data or just Jason-3. b) Right panel shows the r.m.s. difference of the Jason-2 and Jason-3 mean fields as a function of mean conditions. (Data used for this are from Dec. 2016 - Feb. 2017).

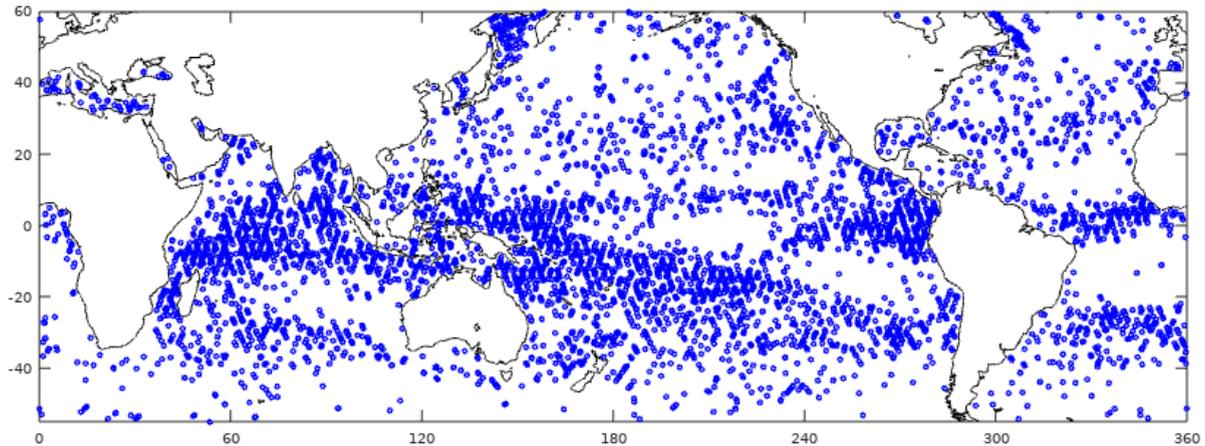
The difference in the gridded fields (Fig. 14a) matches the pattern of the mean  $H_s$  conditions (not shown), with the r.m.s. difference rising monotonically with wave height. This gridding error would be reduced by compiling data over larger grid boxes. Considering Jason-2 and Jason-3 to have near-identical performance, the uncertainty in their individual climatologies would be characterised by the curve in Fig. 14b divided by  $\sqrt{2}$ . Provided all altimeter estimates of  $H_s$  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.

#### 4.2 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. 13b 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 15 shows that the locations of sharp changes in derived  $H_s$  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 15 : Locations where  $\sigma_{H_s} > 1.5m$  (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  $H_s$  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 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. 16 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.

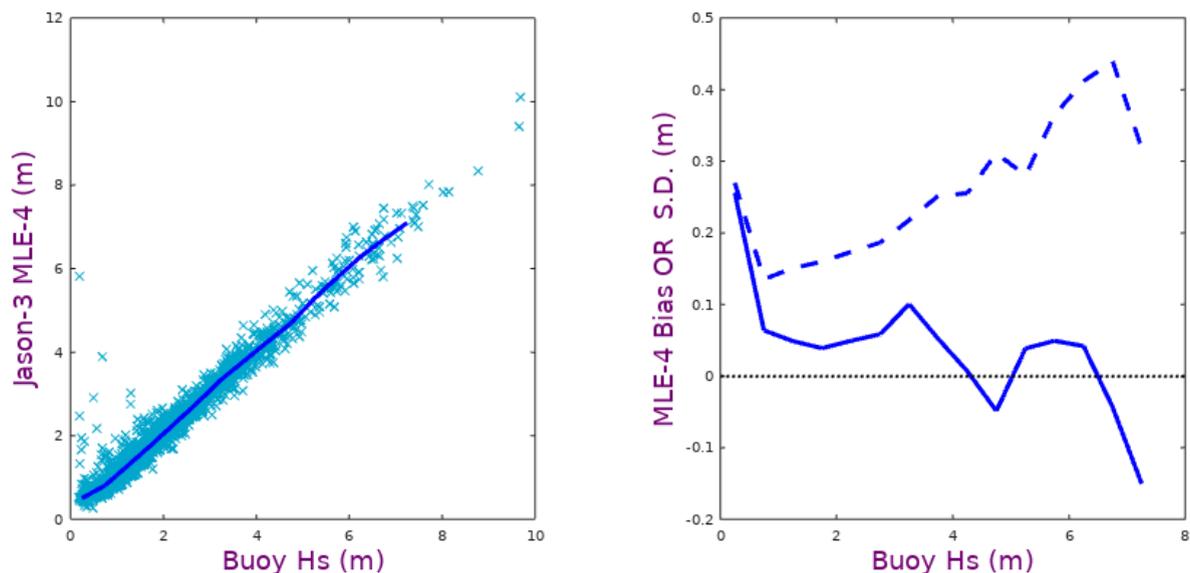


Figure 16 : a) Left panel shows scatterplot of matched up buoy and Jason-3 measurements. The buoy data have been smoothed over 3 hours and then linearly interpolated to the time of the overpass, whilst the altimeter values are the mean of the 51 20 Hz records nearest to the buoy. b) Right panel shows 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 17 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.

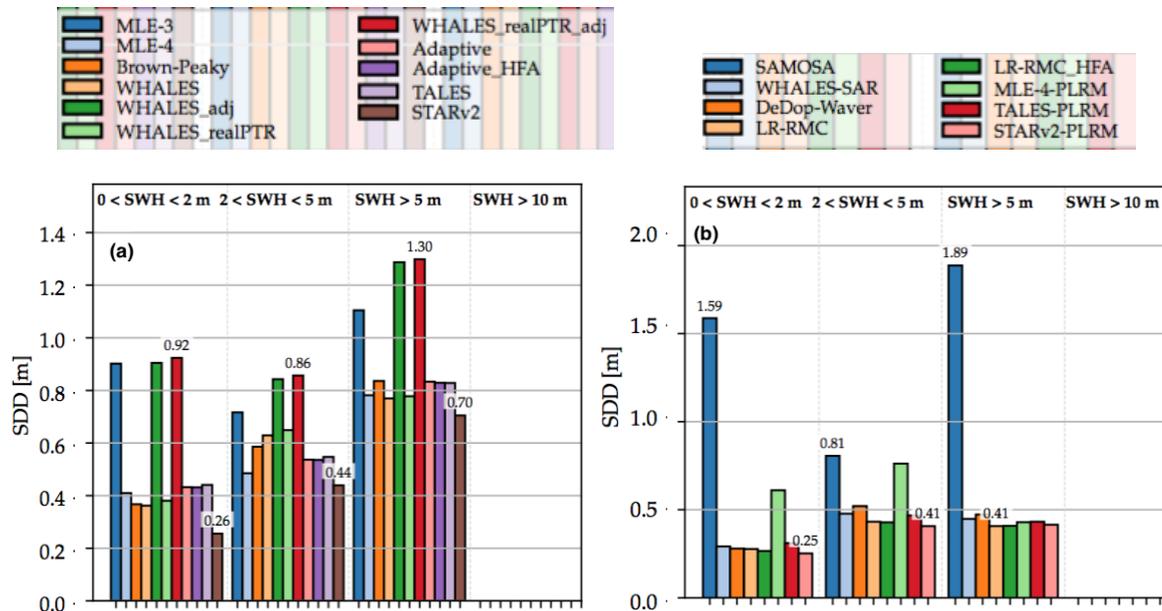


Figure 17 : Standard deviation of differences of various altimeter retrackerers against in-situ buoy data as a function of SWH for a) Jason-3, b) Sentinel-3A respectively. [Taken from Schlembach et al. (2020)].

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. 18a)

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. 19a). 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.

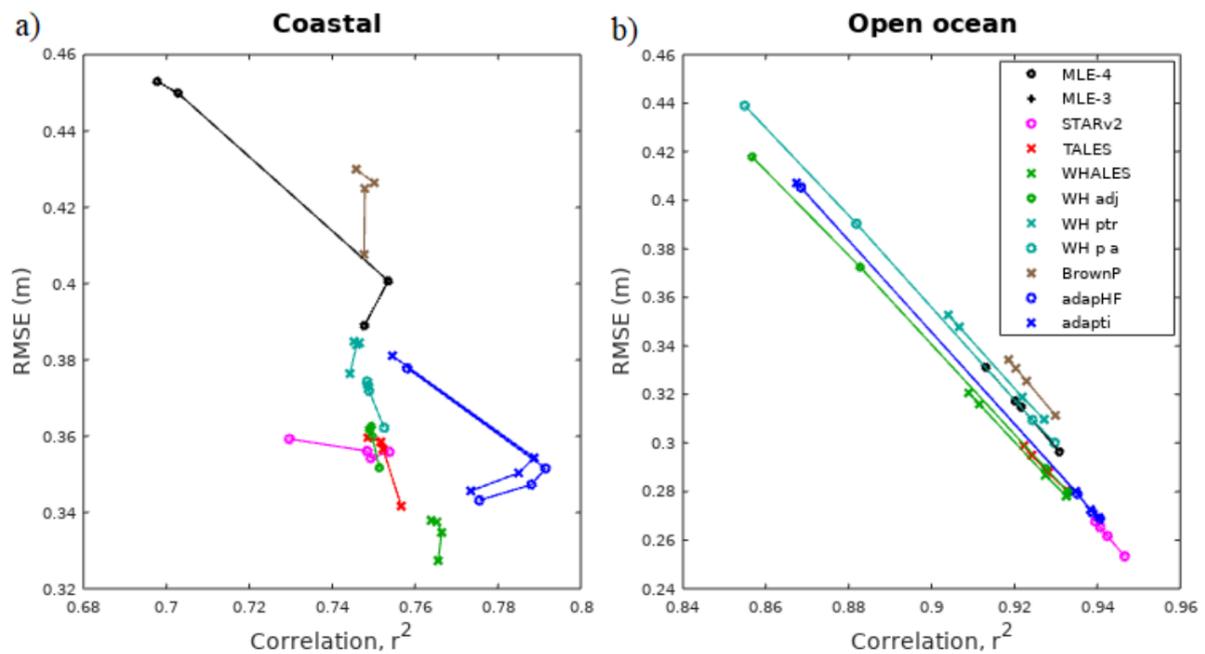


Figure 18 : Generally improved performance as minimum no. of valid measurements is increased from 1 (at top left) with low  $r^2$  and high RMSR) to 5 to 10 to 20. Results are for various retracker applied to Jason-3. [Taken from Quartly & Kurekin (2020)].

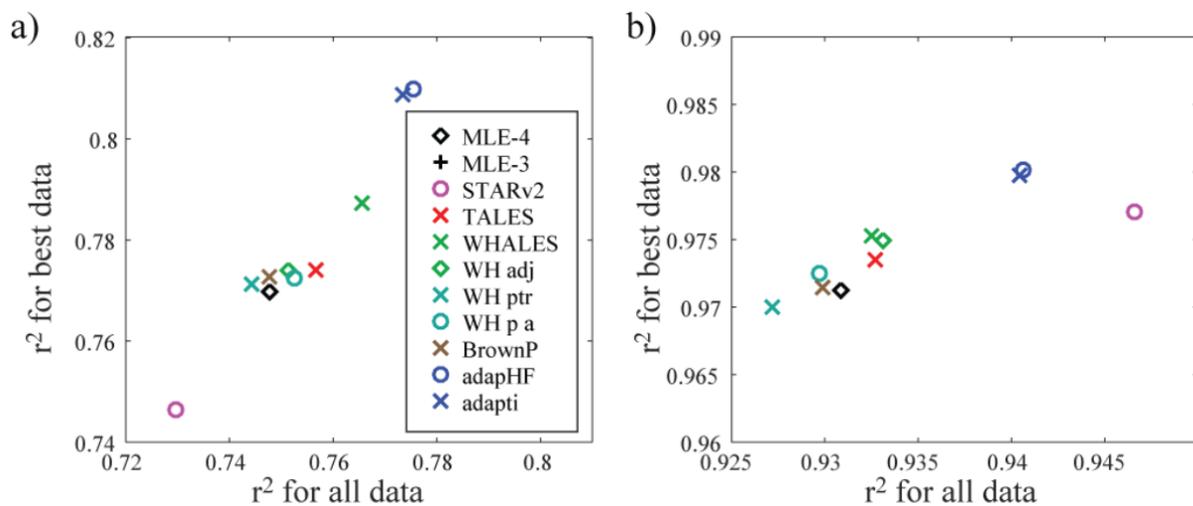


Figure 19 : Effect of match up selection on  $r^2$  metric for (a) Coastal, (b) Open ocean. Results are for various retracker applied to Jason-3; in all cases the  $r^2$  value is higher (better match) when the comparison is restricted to the best aligned buoy-altimeter pairings.. [Taken from Quartly & Kurekin (2020)].

The SAR comparisons with buoys were conducted for 57 buoys found to be collocated with S1 WaveMode imagerettes with a distance less than 50 km to imagerette borders (see Fig. 20). 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. 21).

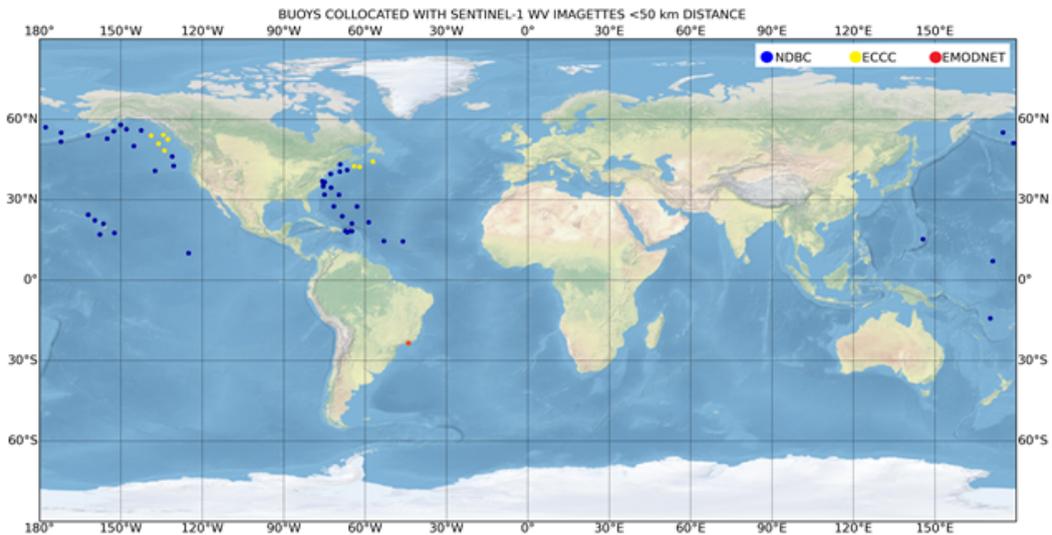


Figure 20 : 57 buoys collocated with S1 WV imagettes with distance less than 50 km to the imagette borders. There are 48 NDBC (blue), 8 ECCC (yellow) and 1 EMODNET (red) buoys.

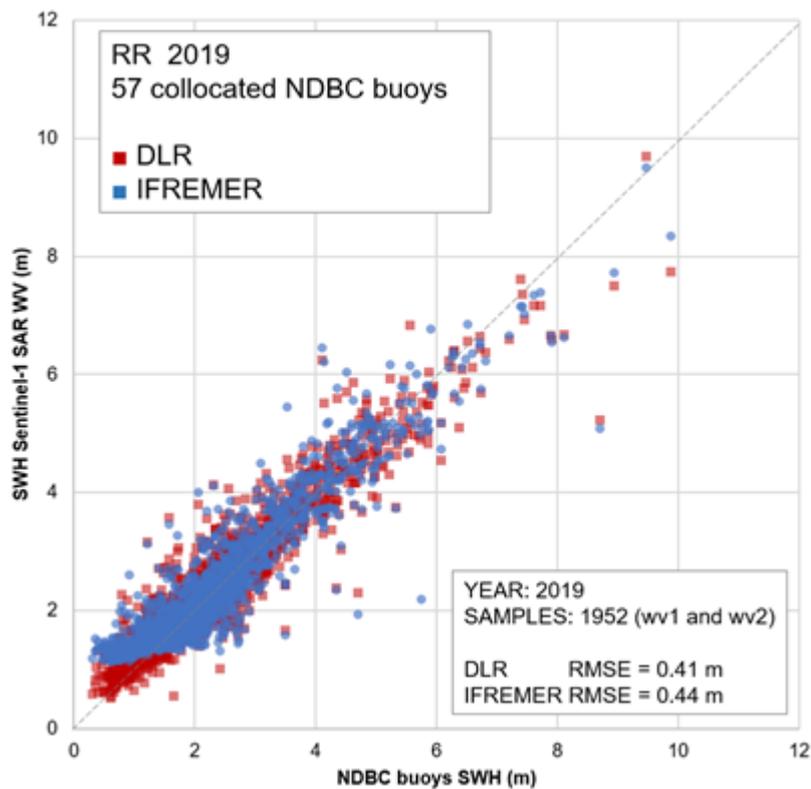


Figure 21 : Scatter plot for SWH comparisons with NDBC measurements conducted for 2019. Averaged RMSE~0.42 m for total SWH.

## 5.2 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. 22) 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.]

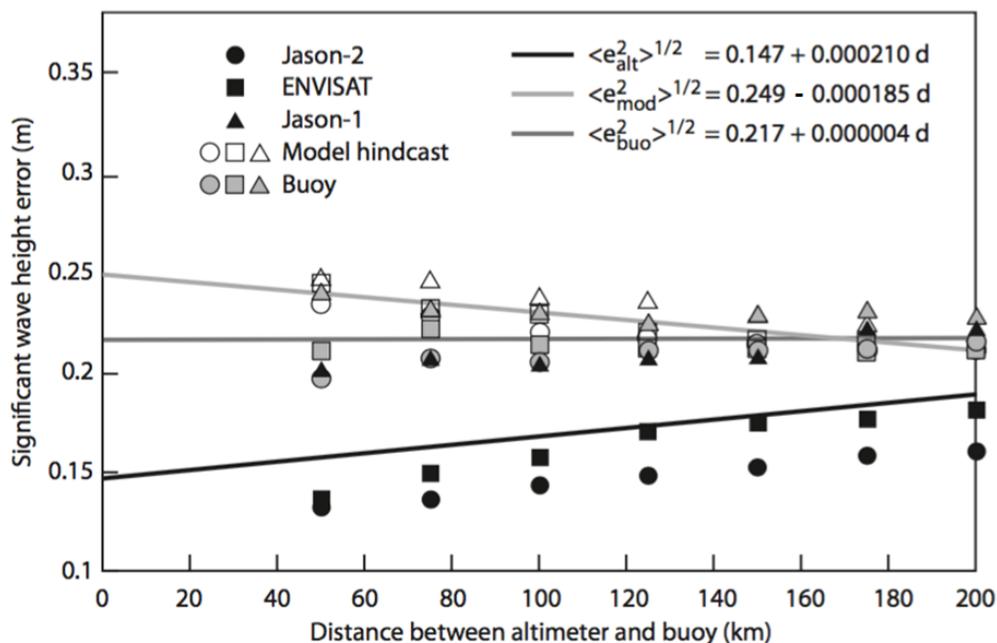


Figure 22 : The partition of errors from a Triple Collocation analysis, indicating that slightly greater errors are attributable to the altimeter as its separation from the buoys increases. [Taken from Abdalla et al. (2011)].

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.

### 5.3 Insight from Climate Assessment Report

#### 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. 23).

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.**

#### 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 1cm/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.

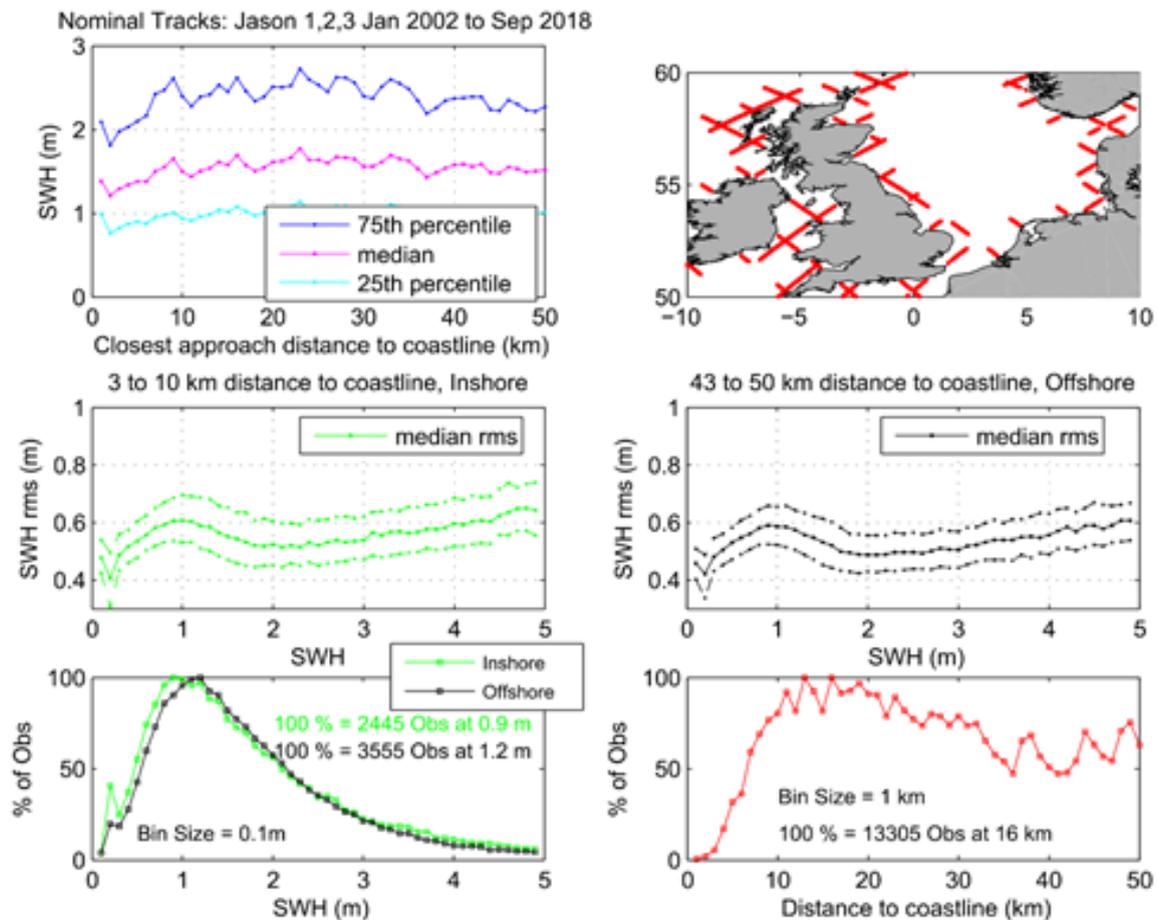


Figure 23 Results from the CCI Sea State L2P product v1.1.using Jason 1, 2 and 3 for the United Kingdom and North Sea region. The SWH as a function of distance to the coast (top left panel), the location of the Jason tracks (top right panel). Inshore waters (3-10 km, middle left panel) and offshore water (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.

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.

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.

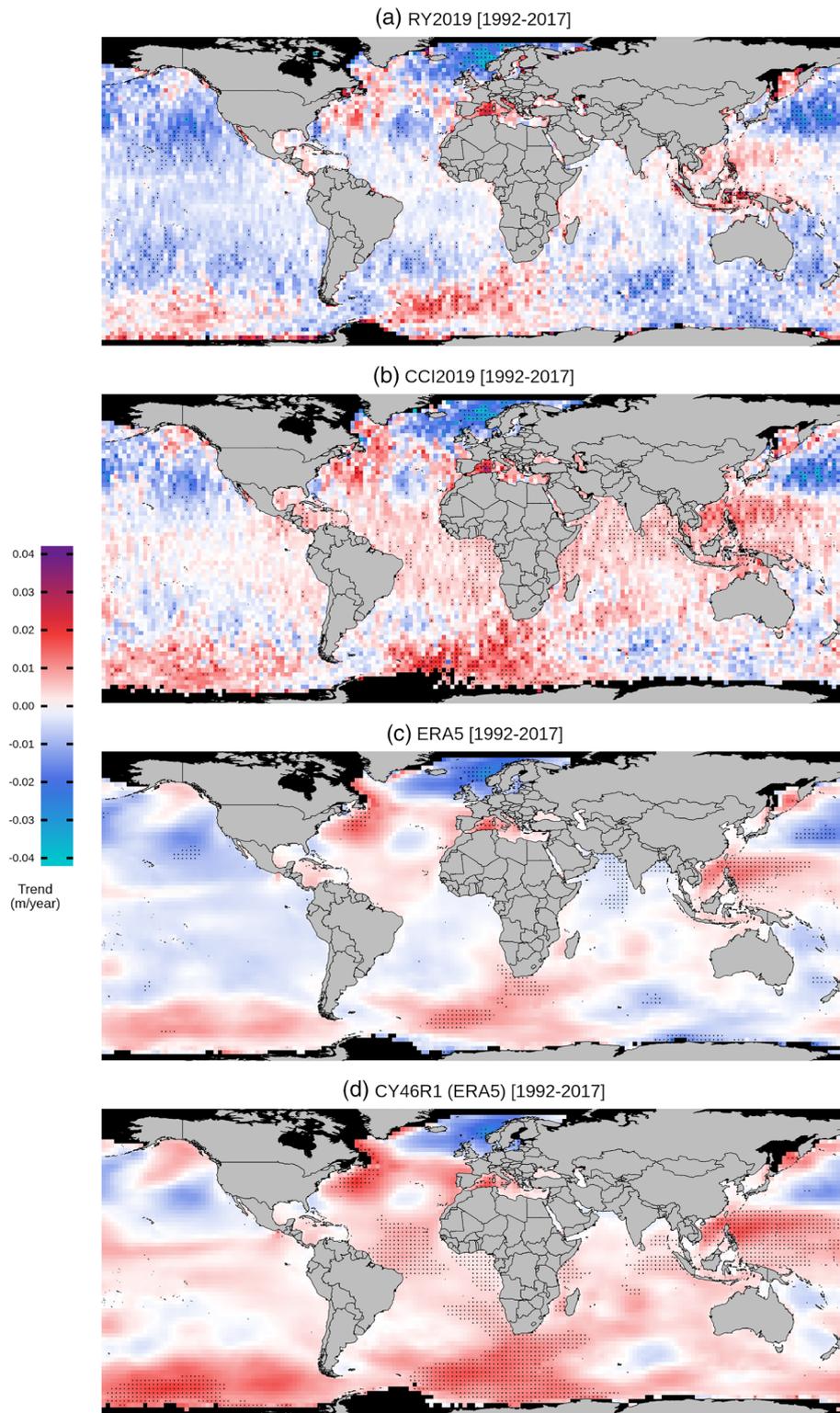


Figure 24 . Global distribution of JFM mean  $H_s$  trend estimates on a  $2^\circ \times 2^\circ$  grid over 1992–2017 for (a) RY2019, (b) CCI2019, (c) ERA5, and (d) CY46R1. Dots indicate grid cells where the trend coefficient is significant at the 95% level. [From Timmermans et al., 2020].

## 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. Also estimates of  $H_s$  from SAR altimetry may 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 0.25-0.35 m	For Jason-3 MLE-4; can be improved with different retracers For 'adaptive' (LRM) and 'LR-RMC' (DDA) retracking algorithms evaluated in Round Robin. Note this noise term increases with $H_s$ , with the increase being faster for DDA retracers than for LRM ones (see Figs. 1,2)
Fading noise on 1 Hz mean	0.12-0.20 m 0,04-0.08 m	For Jason-3 MLE-4. Up to 50% less for AltiKa. This can be improved with different algorithms (see Fig. 6) For 'adaptive' (LRM) and 'LR-RMC' (DDA) algorithms evaluated in Round Robin
Non-Gaussian PTR	$\pm 0.2$ m	But a significant part of this can be corrected for (see Fig. 3)
SAR altimeter imaging of swell	0.1 m	The bias of SAR algorithms w.r.t. LRM ones may depend upon period and direction (see Fig. 5) or simply with $H_s$ (Fig. 6e)
SWH error from SAR	0.25 m $\pm 0.01$ m	S1 WV averaged for both algorithms (DLR and IFREMER), mean of $wv1$ and $wv2$ , validated with CMEMS.
Wave direction error from SAR		no directional parameters provided by SAR algorithm's output.
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. 9b)
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. 11). Mismatch error in Round Robin, where almost all buoys are used (see Fig. 12) but significantly lower errors are achieved with selection of best pairings of buoys and tracks (Fig. 14).
Trends in L4 data)	~ 0.01 m/yr	Fig. 24 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).

### 6.1 Recommended Error Specification in CCI v2 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 v2 product, the characteristics listed here are for WHALES processing of LRM data and LR-RMC processing of SAR mode (delay-Doppler) 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) and  $0.25m + 0.1 Hs$  (SAR, S3A). In both cases 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 of 20 Hz divided by  $\sqrt{19}$  -- or  $\sqrt{39}$ , 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).

There is at present a small but clear difference between values from **LRM and SAR** mode instruments, which will require further R&D to resolve. The estimated "technological error" =  $0.06m + 0.025 Hs$  (Fig. 6f), which could be assigned to the SAR altimeters for now, pending further investigation.

Lastly there is a girding error in producing a **L4 climatology**, which is purely associated with the spatio-temporal sampling. The analysis in section 4 was for  $2.5^\circ \times 2.5^\circ$  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  $[0.15\text{m} + 0.12\text{Hs}] \cdot \sqrt{(3/N)}$ , 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  $0.26\text{ m} \pm 0.01\text{ m}$  dependent on algorithm and data. The SWH noise distribution can be approximated by  $0.20\text{ m} + 0.04 \cdot \text{SWH}$  (see Fig. 25).

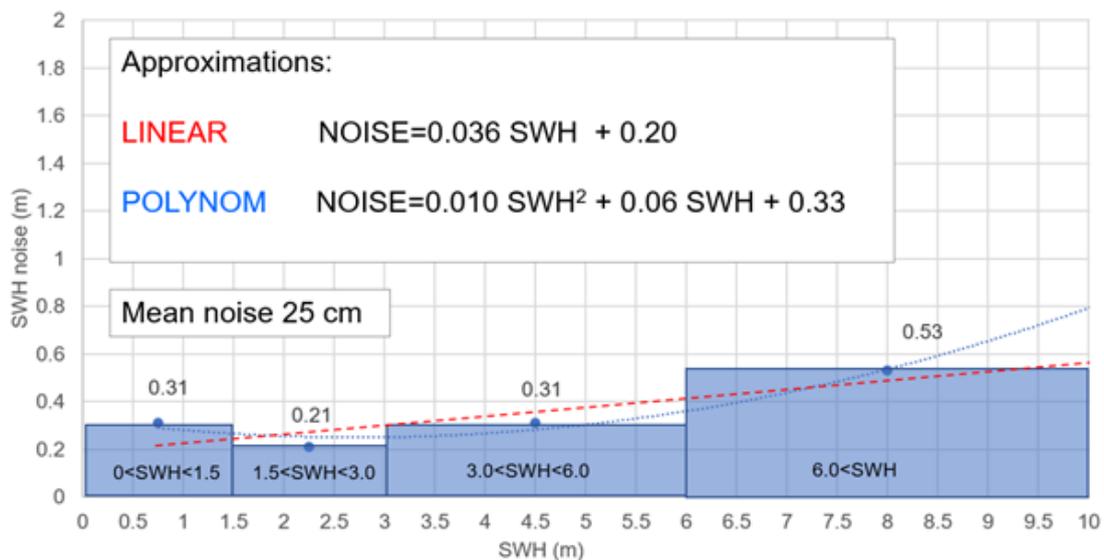


Figure 25 : SAR noise distribution with ground truth using CMEMS. Linear approximation  $\text{NOISE} = 0.20 + 0.04 \cdot \text{SWH}$ .

## References

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.

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 JF , Dodet. G and E. Ash: Sea State CCI Product User Guide v 1.0 , 17 January 2020

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

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.; Passaro, M.; Quartly, G.D.; Kurekin, A.; Nencioli, F.; Dodet, G.; Piollé, J.-F.; Arduin, F.; Bidlot, J.; Schwatke, C.; Seitz, F.; Cipollini, P.; Donlon, C, 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

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.

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