End-to-end ECV Uncertainty Budget (E3UB)

version 1.1, 20 September 2019

Contents

Author	**Approved**	**Signature**	**Date**
Graham Quartly and contributors	Fabrice Ardhuin, Ellis Ash		20 September 2019
ESA Acceptance	Digitally signed by Craig Donlon Date: 2020.02.19 15:57:43 +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

List of Acronyms

ADP	Algorithm Development Plan
ATBD	Algorithm Theoretical Basis Document
cci	Climate Change Initiative
DD	Delay-Doppler
E3UB	End-to-End ECV Uncertainty Budget
ECV	Essential Climate Variable
GDR	Geophysical Data Record
L4	Level 4
LRM	Low Rate Measurement
LUT	Look-Up Table
MLE	Maximum Likelihood Estimator
NRCS	Normalized Radar Cross-Section
PLRM	Pseudo Low Rate Measurement
PTR	Point Target Response
PVASR	Product Validation and Algorithm Selection Report
R.m.s.	Root mean square
RMSE	Root mean square error
S3A	Sentinel-3A
S3B	Sentinel-3B
SAR	Synthetic Aperture Radar
S.D.	Standard Deviation
WHALES	Wave Height Adaptive Leading Edge Subwaveform (retracker)
w.r.t	with respect to
WV	Wave (mode for SAR)

1. Introduction

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

The calculation of significant wave height 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 first version of the End-to-End ECV Uncertainty Budget (E3UB) summarises the state of the knowledge in the first year 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).

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 source of noise 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). Figure 1a shows σ_{Hs} as a function of Hs for Jason-3.

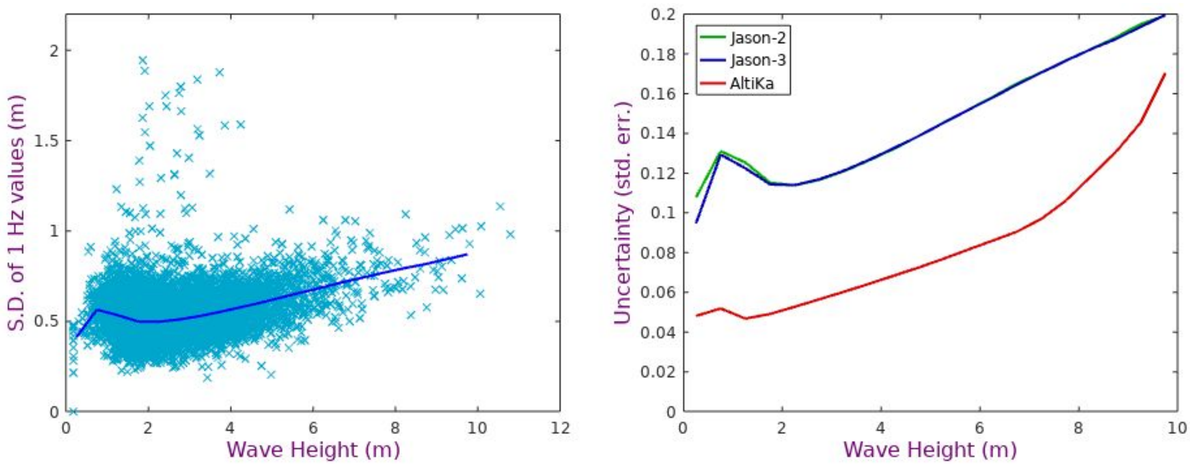

Figure 1 : a) Left panel shows a scatter plot of σ_{Hs} as a function of Hs for Jason-3's MLE-4-based estimates during Jan. 2017. The cyan crosses only show a small subset of the points, but the blue line shows the mean relationship derived from all data (averaged in 0.5m bins). b) Right panel shows 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 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. 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.

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 2a shows the inferred instrument correction on Jason-3 to compensate for the real PTR shape.

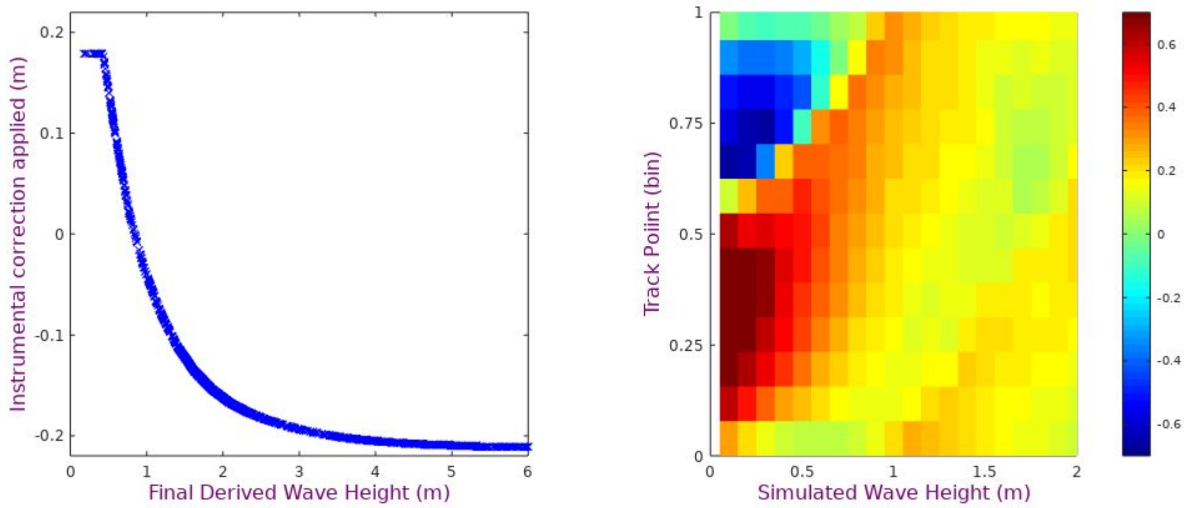

Figure 2 : a) Left panel shows LUT correction for MLE-4 algorithm applied to Jason-3. b) Right panel shows mean bias observed using simulation results for the WHALES algorithm, with the bias varying according to both wave height and the position of the leading edge.

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, based on the values shown in Fig. 2 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.

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 and trailing edges 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. 3a) is similar to that for LRM. This may improve significantly with the development of further SAR retrackers and their tuning.

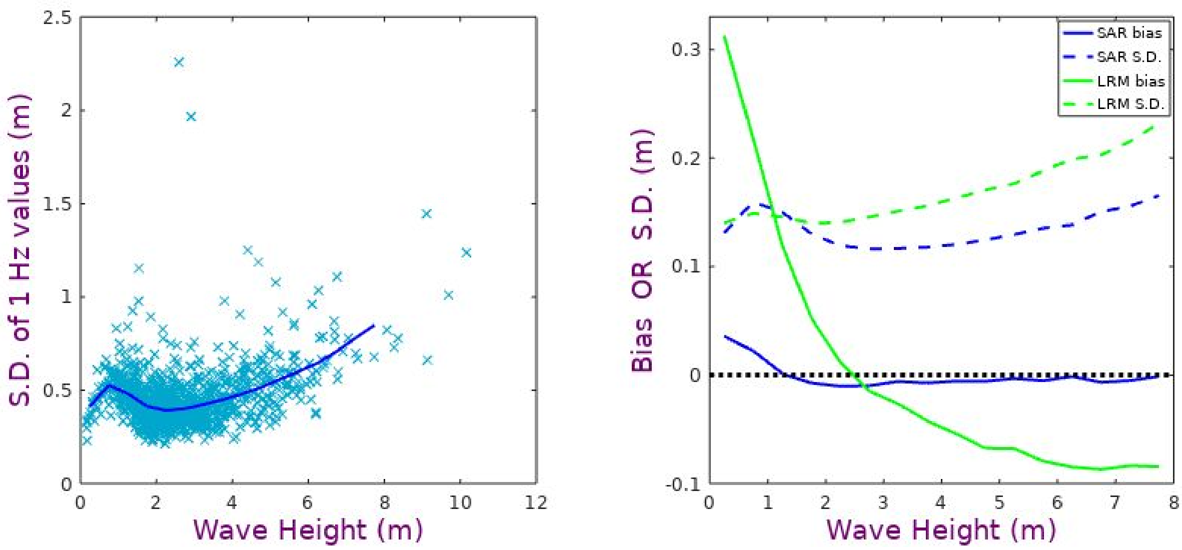

Figure 3 : a) Left panel shows σ_{Hs} (S.D. of values in a 1-second interval) as a function of mean Hs. Data are from Sentinel-3A cycle 033. Cyan crosses show a subset of the points, the blue line shows the mean relationship derived from all the data in the cycle. b) A comparison of S3B data with near-simultaneous 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 0.15m (Fig. 3b); however with S3B in LRM mode its bias relative to S3A varies between 0.3m 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. 4a). 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. 4b).

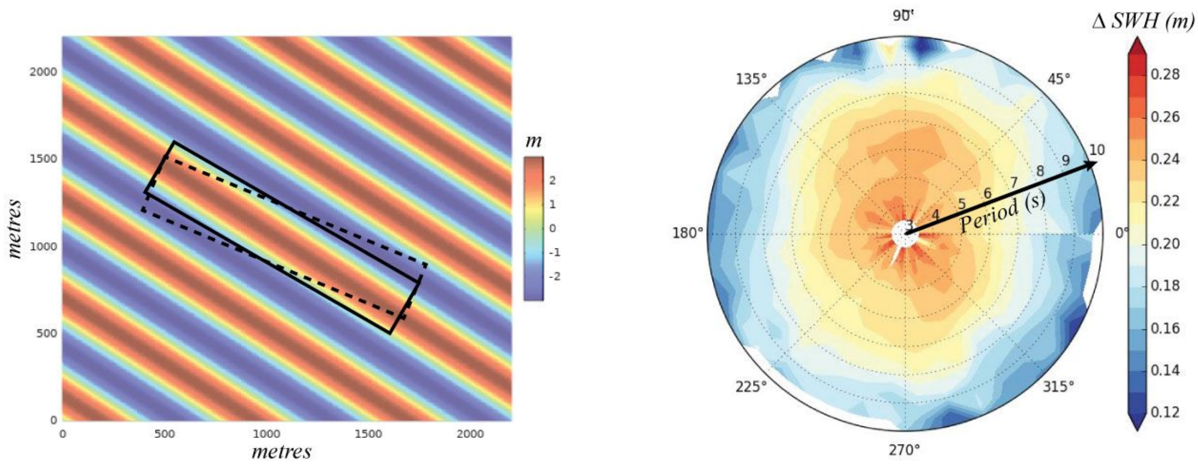

Figure 4 : a) Left panel shows schematic of a simple long wavelength swell field (colours indicating the height of the water surface), with the narrow SAR mode footprint almost aligned with direction of swell. b) Difference between SAR and PLRM estimates of wave height as a function of angle between 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).

2.3 Synthetic Aperture Radar

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 with the exception of cases of rain or snow, directly changing the sea surface structure.

When the algorithms for SAR sea state retrieval are tuned by utilizing models and buoys as ground truth, then uncertainties of both ground truths are propagated into the resulting algorithm's accuracy. Worldwide, model data (e.g. WWIII) and buoys agree with RMSE~25 cm for general conditions which, however, can reach RMSE~50 cm under storm conditions, mostly due to a relatively small temporary shift of the storm peak passing through a buoy position.

However, environmental and man-made conditions are significant: e.g. under strong winds, the sea surface can be completely destroyed by wave breaking; and man-made artefacts like ships, ship wakes, or wind farms can spoil the sea state signal.

2.4 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 that may differ by half an hour from the time of the altimeter overpass, and iii) there are errors in the buoy measurement, such as them not being able to sample the whole wave spectrum. 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. Thus the illustration in Fig. 5 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.

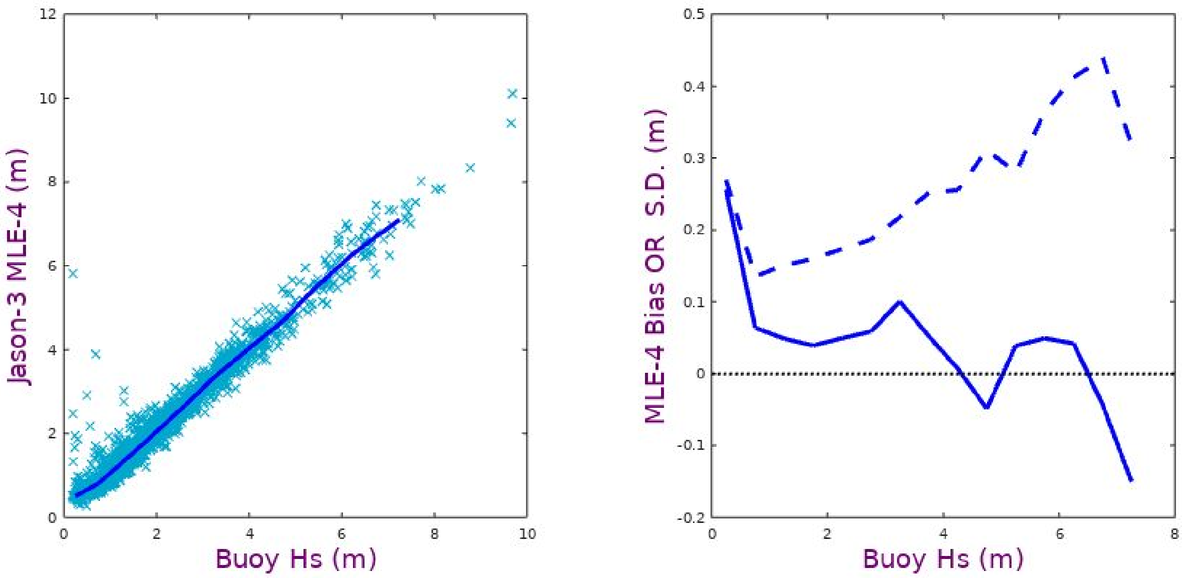

Figure 5 : a) Left panel shows scatterplot of matched up buoy and Jason-3 measurements. The buoy data have been smoothed over 3 hours and then linearly interpolated to the time of the overpass, whilst the altimeter values are the mean of the 51 20 Hz records nearest to the buoy. b) Right panel shows 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.

For comparisons of SAR data with buoys, there is similarly a source of uncertainty resulting from the spatial variability in sea state fields by differences in SAR scene location and buoy location. Imagettes from Sentinel-1 WV mode provide estimates for a covered region typically 5 times a month, with an average collocation displacement (for the NOAA buoys) of ~20 km. A simple comparison of model results at SAR scene location and at buoy location at the scene acquisition time results in a RMSE of ~0.20 m. This uncertainty does not exist with other SAR acquisition modes, as these cover larger areas and often include shelf regions, which are rich with buoys.

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). Thus 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. 6, 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.

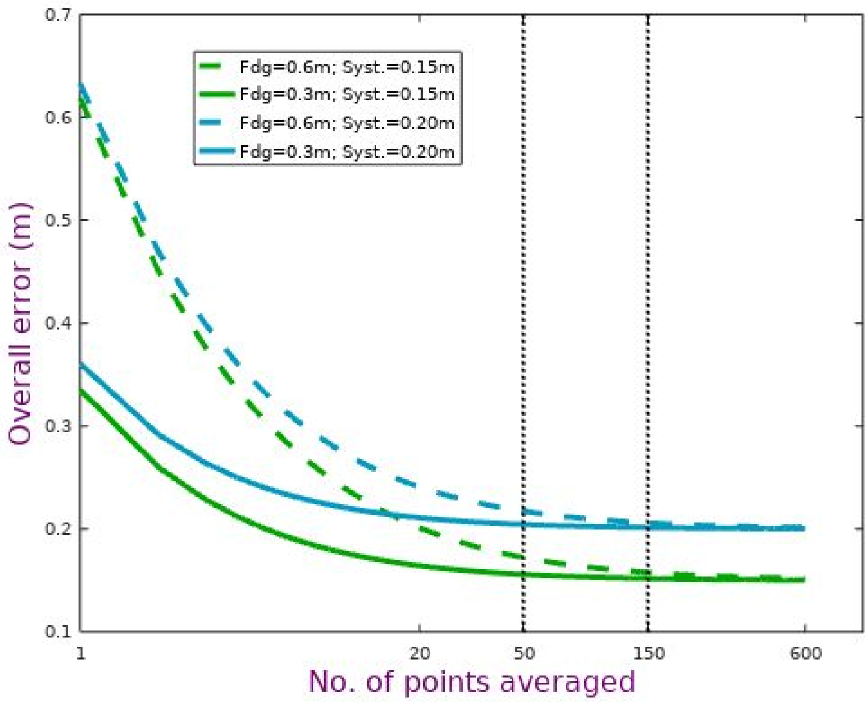

Figure 6 : Illustration to show that error due to fading noise dominates for scales of 1 Hz (20 obs.) and finer, but systematic errors are more important at larger scales. Dashed line at 50 pts represents the averaging scale used in buoy comparisons, and that at 150 pts equates to 50 km, a typical scale 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 geograpically-correlated errors.

4. Uncertainty in L4 gridded products

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. 7).

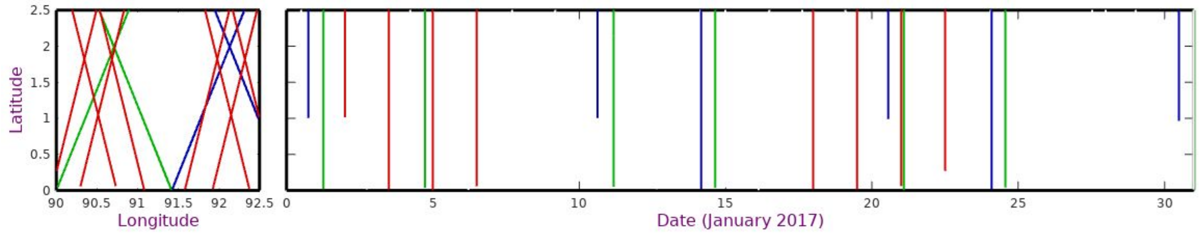

Figure 7 : a) Left panel shows altimeter tracks across a randomly-selected open ocean box for January 2017. Jason-3 (in the reference orbit) is in blue; Jason-2 (in the interleaved orbit) is in green; AltiKa (in its drifting orbit) is in red. [Note AltiKa's track spacing is much finer than when in 35-day 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), 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. 8).

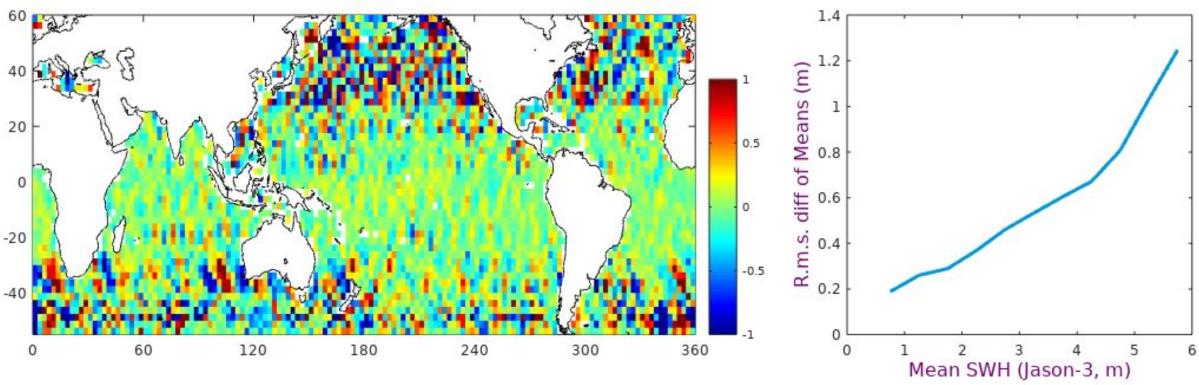

Figure 8 : a) Left panel shows difference in mean wave fields for January 2017 using either just Jason-2 data or just Jason-3. b) Right panel shows the r.m.s. difference of the Jason-2 and Jason-3 mean fields as a function of mean conditions. (Data used for this are from Dec. 2016 - Feb. 2017).

The difference in the gridded fields (Fig. 8a) 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. 8b divided by $\sqrt{2}$. 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.

4.2 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. 7b 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.

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 9 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.

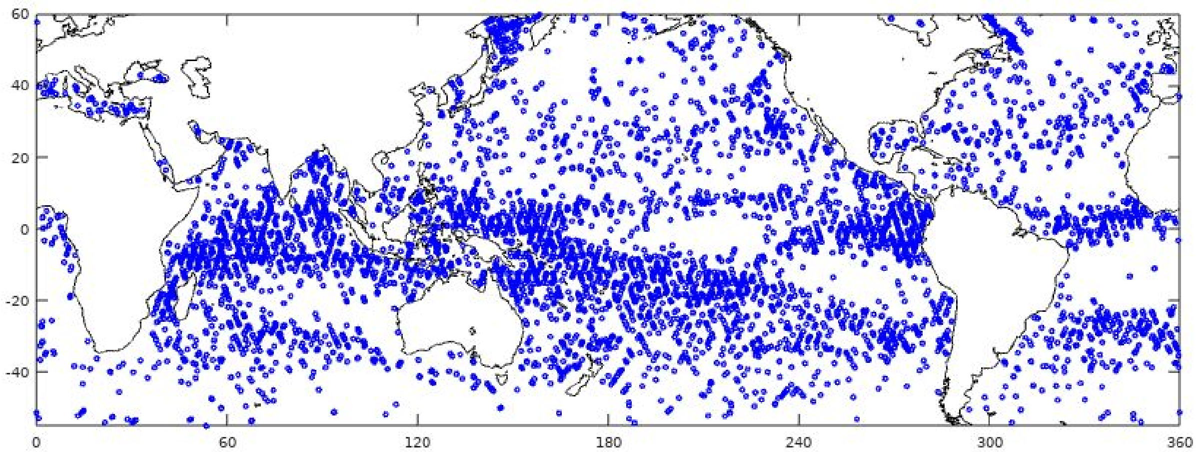

Figure 9 : Locations where σ_{Hs} >1.5m (see Fig. 3a for mean values). [Note this indicates occasional high values for Jason-3 during Dec. 2016-Feb. 2017, rather than the typical values there].

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. 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 Hs 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 in the limitations of the sampling pattern. 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 algorithms
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. 6)
Non-Gaussian PTR	±0.2 m	But a significant part of this can be corrected for
R.m.s. error for ~17 km section, compared with best buoys	0.15 m	This also includes the errors in the buoys and the difference in space and time sampling
SAR altimeter imaging of swell	0.1 m	Difference w.r.t. LRM depends upon period and direction
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.