Article

Sensitivity of Altimeter Wave Height Assessment to Data Selection

Graham D. Quartly * and **Andrey A. Kurekin**

Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, Devon PL1 3DH, UK; anku@pml.ac.uk
* Correspondence: gqu@pml.ac.uk

check for updates

Abstract: This paper addresses the issue of how the selection of buoys and the calculation of altimeter averages affect the metrics characterising the errors of altimetric wave height estimates. The use of a 51-point median reduces the sensitivity to occasional outliers, but the quality of this measure can be improved by demanding that there is a minimum number of valid measurements. This had a marked impact in both the open ocean and the coastal zone. It also affected the relative ordering of algorithms' performances, as some fared poorly when a representative value was gleaned from a single waveform inversion, but had a much better ranking when a minimum of 20 values were used. Validation procedures could also be improved by choosing altimeter-buoy pairings that showed a good consistency. This paper demonstrated an innovative procedure using the median of the different retrackers analysed, which can be easily extended to other data validation exercises. This led to improved comparison statistics for all algorithms in the open ocean, with many showing errors less than 0.2 m, but there was only one strong change in the relative performance of the 11 Jason-3 retrackers. For Sentinel-3A, removing the inconsistent coastal buoys showed that all of the new algorithms had similar errors of just over 0.2 m. Thus, although improvements were found in the procedure used for the Sea State Round Robin exercise, the relative rankings for the buoy calibrations are mostly unaffected.

Keywords: altimeter; validation; buoys; ocean; coastal zone; data-flagging; sensitivity studies; Jason-3; Sentinel-3A

1. Introduction

Significant wave height (SWH) is listed by GCOS (Global Climate Observing System [1]) as an "essential climate variable", since the study of its geographical and temporal variation is an important component in understanding the evolving environment of our planet. In coastal regimes, waves may have benign effects (such as creating good surfing conditions) as well as hazardous consequences (such as damage to moored vessels and harbour structures and inundations in low-lying areas). In deeper waters, waves make a big impact through damage to shipping and offshore structures, but are also important for the mixing they produce both within the surface waters (breaking down the thermocline and bringing nutrients up) and in enhancing air-sea exchange of gases [2].

Satellite-borne radar altimeters are the primary means of providing near-global measurements of SWH, with datasets being collated from multiple instruments being combined to give 33 years of near-continuous coverage that has been used in various climatological studies [3]. In an endeavour to improve the accuracy and consistency of these multi-satellite compendia, the European Space Agency (ESA) has inaugurated a "Sea State" component to their Climate Change Initiative (CCI) programme [4]. One focus of this work has been to instigate a Round Robin process to evaluate a number of candidate algorithms, and assess their reliability in providing data, with low noise levels and minimal outliers, which compare well with model forecasts and the limited coastal buoy network [5].

Naturally, the metrics for the assessment (bias, standard deviation, correlation coefficient) vary with the data selection criteria for the altimetry and the choice of buoys used, according to their degree of exposure, proximity to coast, and wave regime encountered, with many previous studies concentrating the evaluation in open ocean conditions [6–8]. However, given the particular challenges of altimetry close to the coast, namely radar returns from nearby land and the presence of "bright targets" that are associated with sheltered bays [9,10], there has been a strong recent push to develop new technology and algorithms that are more robust in such situations.

In the development of the Round Robin exercise, the specific metrics to be used, the buoy selection, and the processing to be applied to the altimeter products were all published in advance with the aim of transparency about the procedures to be followed [11]. However, subsequent experience suggested that the methodology could be improved, by incorporating further constraints on exactly which data were to be used. In this paper we investigate the repercussions of applying greater initial quality control, and of how that not only affects the measures for all candidates, but changes their ranking. In Section 2, we start with a brief reprise of the buoy and altimeter data being used, and then in Section 3 highlight some of the issues affecting data quality; Section 3.1 shows the sensitivity to the threshold on minimum number of valid observations and Sections 3.2 and 3.3 detail the identification of a consistent set of altimeter-buoy pairings and of how that affects results. The work is brought together with a discussion in Section 4.

2. Datasets

2.1. Buoy Data

We make use of a compilation of buoy and altimeter data gathered for the Sea State CCI Round Robin exercise [5]. The in situ data come from the operational archive at the European Centre for Medium-range Weather Forecasting (ECMWF) augmented by a few sources (see Section 2.2 of [5]). Because only buoys near the selected altimeter tracks were used, this means that only a small fraction of the ECMWF dataset was employed. These instruments provide a lot of information on the wave field, with many also providing sea surface temperature and wind speed and direction; however, only the buoy estimates of significant wave height (SWH) are used in this study, as that is the measure most readily obtainable from satellite altimeter data. As part of the rationale for this study is to investigate altimeter SWH retrieval in the coastal zone, this collated dataset is not restricted to deep ocean well exposed buoys as have been used in some studies [6,8,12]. The in situ data are from a set of buoys of varying design, water depth, exposure, and distance to coast. These "ground truth" data originate from the manufacturers' standard algorithms, without individual recalibration, whereas Challenor and Cotton [13] and Durrant et al. [6] have suggested that adjustments may be needed specific to the observing network. Consequently, there may be inhomogeneities within the reference dataset.

2.2. Altimeter Data

To best enable a fair comparison across multiple algorithms, it was necessary that they should all process the same altimeter data. Because the various data providers would need to work with the detailed altimeter waveform data, which necessitates significant processing time, they were each asked to process a specified subset of Jason-3 and Sentinel-3A tracks for a given two-year period. This guaranteed that evaluations would cover all seasons, but is not a truly global comparison because the majority of the buoys, which are run by the operational weather agencies, are in the northern hemisphere and close to the continents.

The satellite instruments chosen (Jason-3 and Sentinel-3A) are just two of those flying during the validation period (June 2016 to June 2018), and were selected as they occupy orbits likely to provide the backbone of any long-term climatology, and also because these two instruments use different technologies. Jason-3 (J3) is the fourth in a series of satellites to utilise a 10-day repeat orbit sampling the Earth's surface between 66°S and 66°N, with a longitudinal separation of parallel tracks of 2.83°.

It records waveforms ("radar echoes" from the sea surface, see Figure 1a) in the manner of conventional radar altimeters, which is known as "Low Resolution Mode" (LRM), with many algorithms having been developed to extract geophysical estimates from such returns.
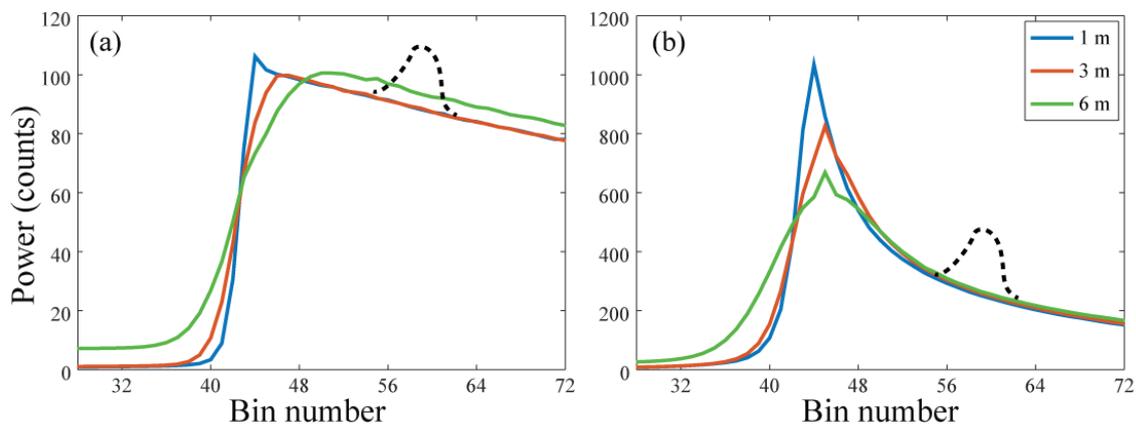
Figure 1. (**a**) Typical Low Resolution Mode (LRM) waveform; (**b**) Typical Doppler Altimeter (DDA) waveform: the coloured lines show the shape for three different significant wave height (SWH) values; the black dashed line shows the potential land contribution near the coast. This will normally move towards the front of the waveform and become more prominent as the coast is approached.

Sentinel-3A (S3A) is in a 27-day repeat orbit, spanning between latitudes 82°S and 82°N with a longitudinal track spacing of 0.94°. It operates a delay-Doppler Altimeter (DDA, otherwise known as "SAR mode"), whereby the extra information in the received signal should enable better focus on the sea surface, with expected gains in footprint size, resilience to land reflections, and an increase in the number of independent pulses that can be averaged [14]. This yields a sharper waveform than for LRM, with both a steeper leading edge and trailing edge (see Figure 1b). The data from Sentinel-3A are also processed without the Doppler information to yield pseudo-LRM (PLRM) waveforms to which conventional trackers can be applied in order to permit an understanding of the relative differences between LRM and DDA processing (an overview of the performance of Sentinel-3A is given in Quartly et al. [15]).

The standard algorithm for parameter estimation from LRM or PLRM waveform data is MLE-4 (Maximum Likelihood Estimation fitting four parameters [16]), which fits a modelled shape to the full extent of the waveform data. From this it infers the four parameters: range, SWH, normalized backscatter, and mispointing, with the first two being most dependent on the shape and position of the leading edge, and the latter two being affected by variations in the trailing edge (an earlier standard algorithm had been MLE-3, which also analyses the whole waveform, but does not fit the mispointing term).

The various candidate algorithms are described in some detail in the Round Robin evaluation [5]; to avoid repetition, only the briefest details are recapped here.

One particular challenge in deriving SWH from LRM instruments has been the performance at very low wave heights (<1 m), where all of the useful information is recorded within a very few bins spanning the leading edge. This can also be exacerbated by uncertainty in the instrument's Point Target Response (PTR), which has usually been modelled by a narrow Gaussian pulse. The other tough challenge has been near-coastal environments where reflections from land or sheltered bays may contaminate the main signal, with the effects being first noted within the trailing edge of the waveform, and moving to the leading edge as the coast was approached.

A number of algorithms have been developed to downweight or totally ignore the power in these later waveform bins in order to increase the robustness in the coastal zone. This was originally pioneered by ALES (Adaptive Leading Edge Subwaveform) retracker [17], which had a focus on improving range information. In this paper, we consider a number of variants on this approach:

(i) WHALES [18], which is a version of ALES optimised for Wave Height, (ii) TALES [19], which uses a numerical model for the retracker, and (iii) Brown-Peaky [20,21], which codes a decision for whether an extraneous peak exists and, only in those cases significantly downweights that part of the waveform. There was also a version of WHALES, which included a novel correction for the PTR [18]. A somewhat different approach is taken by the *adaptive* retracker, which first performs a classification of waveforms according to shape [22], fits a four-parameter shape model (where the fourth term represents anisotropy [18]) and solves this using a numerical representation of the PTR rather than a Gaussian approximation.

All of the above have individually treated each waveform; however it may be expected that successive estimates, based on largely overlapping instrument footprints, should have little change between them. Sandwell and Smith [23] had shown there is significant cross-talk between the errors in range and SWH, as both are derived from the leading edge; this can be quantified and utilised to reduce the noise in SWH [24]. This "high-frequency adjustment" was applied to output from the WHALES, WHALES_PTR and adaptive retrackers. It is not actually a smoothing of the SWH estimates, but application of a correction that involves utilising input from neighbouring waveforms. Generally, the effect is to remove the high-frequency measurement noise, but leave the local average unaffected and, thus, these variants will show similar mean values to their original retracker output. A markedly different approach is adopted by STARv2 [25], which identifies and retracks many individual subwaveforms, and then seeks a consistent trajectory through the "point cloud" of solutions. In effect, this leads to a very smooth SWH profile along-track, but with the level of smoothing inherent to the approach rather than applied afterwards.

For DDA waveforms, changes in SWH affect both the leading and trailing edges of the shape (see Figure 1b). As the Doppler operation better constrains the position on the sea surface being viewed, the spurious extra peaks on the trailing edge associated with land or bright target contamination are not so significant for DDA instruments. Thus, a greater number of waveform bins can be utilised, potentially improving the robustness of algorithms to the effect of fading noise. The complicated mathematical model for the shape is named "SAMOSA" (SAR Altimetry MOde Studies and Applications), with the standard inversion algorithm usually referred to by the same name [26].

WHALES-SAR adapts the WHALES approach i.e., identifying the subwaveform segment that contains the leading edge and then applies the Brown retracking model to that. As the DDA waveform has a different shape, there is then an empirical non-linear rescaling in order to provide the SWH estimate [18]. The LR-RMC method applied Range Migration Correction and Compression separately to each group of four bursts and averages the beams to generate a 20 Hz waveform product. This, in effect, creates an instrument footprint that is similar in size to that of LRM products. It is stated to overcome the perceived problem of SAMOSA in dealing with the effect of swell [27]. As for the *adaptive* algorithm, a numerical retracking approach is used, which can incorporate the true instrument PTR. DeDop-Waver implements a reprocessing of Level 1A data to Level 1B, equalising the contributions of the different Doppler beams, which is followed by the application of the Makhoul et al. [28] ocean retracker to the radar scattering model of Ray et al. [26].

More detailed information on all of these algorithms is given in the earlier paper [5] and in the *Algorithm Theoretical Basis Document* [18]. They are summarised in Table 1.

2.3. Data Match-Up

Because current altimeters only provide data at nadir, they do not usually sample the exact locations where in situ measurements are made by buoys. Typically, for open ocean applications, researchers consider buoy data within 50 km and 30 min of the altimeter measurement to be appropriate. These values were determined by Monaldo [29] based on examination and auto-correlation analysis of altimeters and buoys, respectively. Nencioli and Quartly [30] showed that aspects of the coastline

morphology and orientation to the prevailing waves can be important in the coastal zone, and this is an issue that we address later.

Table 1. Summary of the different retrackers investigated. Retracker 3 selects a smooth profile, and retrackers 6, 8, 11, & 16 include smoothed corrections applied to 5, 7, 10, & 15, respectively, and they are indicated by round plot symbols in some subsequent plots.

	Short Name	Waveform Type	Data Source	Brief Description
1	MLE-4	LRM	J3 , S3A	Maximum Likelihood Estimator for 4 unknowns
2	MLE-3	LRM	J3	Maximum Likelihood Estimator for 3 unknowns
3	STARv2	LRM	J3 , S3A	Finds best route through a cloud of subwaveform solutions
4	TALES	LRM	J3 , S3A	TU-Darmstadt version of ALES retracker
5	WHALES	LRM	J3	Wave Height optimised version of ALES retracker
6	WH adj	LRM	J3	WHALES output adjusted for variations in estimated range
7	WH ptr	LRM	J3	WHALES output corrected for real PTR
8	WH p aj	LRM	J3	WHALES output corrected for real PTR and variations in estimated range
9	BrownP	LRM	J3	Brown-Peaky: discards data around spurious peaks and fits Brown model
10	adapti	LRM	J3	Adaptive waveform fitting, using numerical description of real PTR
11	adapHF	LRM	J3	adapti adjusted for variations in estimated range
12	SAMOSA	SAR	S3A	Default algorithm for S3A DDA waveforms
13	WH-SAR	SAR	S3A	Adaption of WHALES approach for SAR waveforms
14	DWaver	SAR	S3A	Implementation of Level 1B processing using DeDop toolbox
15	LR-RMC	SAR	S3A	Range Migration Correction and Compression + numerical retracking
16	LR-RHF	SAR	S3A	LR-RMC output adjusted for variations in estimated range

In this paper, we re-use the match-up database provided for the Round Robin paper [5], so the information is only recapped here. Jason-3 and Sentinel-3A both provide waveforms and, thus, opportunity for SWH estimates at a rate of ∼20 Hz, corresponding to an along-track displacement of ∼294 m (J3) or ∼334 m (S3A). Having identified which buoys were close to which Jason-3 and Sentinel-3A tracks, an automatic procedure was used to find the point of nearest approach and extract the SWH estimates for ±25 points along-track i.e., a total of 51 high-rate measurements, corresponding to 15–17 km along-track. The buoys stored data hourly or every half hour, and these were smoothed in time with a three-point running mean filter. These databases itemised 7128 buoy measurements coinciding with Jason-3, with those coming from 135 different buoys. As there were more different S3A tracks than for J3, but with far fewer repeats, there were 2535 buoy observations for the S3A comparison, but coming from 191 different buoys.

3. Factors Affecting Comparisons

There have been many papers looking at the accuracy of altimeter estimates of SWH, and the values differ depending upon the instrument, the retracking algorithm applied, and the choice of buoy data used for match-ups. For buoys far from the shore, the typical match-up criteria are to use altimeter data within 50 km of the buoy, and buoy data within half an hour of the altimeter pass. For this, there are underlying assumptions of spatial and temporal homogeneity; but, even in the deep ocean, currents or eddies may significantly modify the wave field. However, one of the aims of the Round Robin assessment conducted by Schlembach et al. [5] was to evaluate a new set of algorithms in the coastal zone. The distance from coast at which land will still affect the estimation will depend upon the topography of the land, how reflective it is, and the sensitivity of the algorithm to spurious reflections.

As the effects of land and sheltered bays become more apparent in the waveforms (Figure 1), the algorithms are less able to cope and either return spurious values or simply flag the data as 'invalid'. Figure 2 shows how the data return rate for a number of different retrackers falls as the coast is approached. The proportion of invalid data for a particular algorithm will depend upon the data provider's implementation of a quality flag, which may have a threshold that is adjustable for different applications. The plots presented in Figure 2 illustrate the variation in the data providers' implementations of a quality flag. What is key here is that most retrackers show significant data loss by 5 km from the coast, and those that show a high data return closer in may just be failing to flag the

problems that exist. Of course, the presence of data does not imply that they are necessarily useful; the range to coast at which data will be affected will depend not only on the sensitivity of the retracking algorithm, but also the height of the land and how reflective it is. A targetted comparison with buoys around southwest UK showed increased errors (reduced accuracy) for MLE-4 in the nearest 15 km [30], whilst an analysis of that algorithm around Australia showed the 1 Hz records to deteriorate at about 10 km from the coast [21].
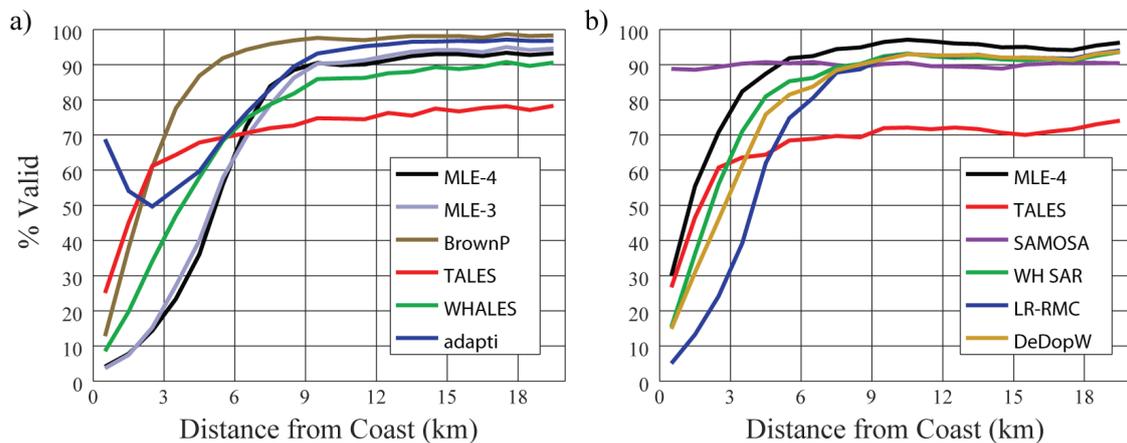
Figure 2. Data return rates for various retrackers as the altimeters approach the coast. Plots show percentage flagged as acceptable estimate for (**a**) Jason-3 and (**b**) Sentinel-3A. The data here are "quasi-global" i.e., from the two years of selected tracks used for the Round Robin, but avoiding areas of sea-ice, and are binned at intervals of 1 km.

Secondly it is a requirement that the buoy data be accurate, or at least consistent. However, there have been many evolutions in the design of buoys, with small spherical ones and a range of larger toroidal ones plus spar buoys, with these different designs especially affecting the range of wave periods to which they can respond. Calibrations may be available for each individual buoy, but these are not always available at the time data are circulated, so that measurements from some buoys may be biased.

Thirdly, there is the question of whether the altimeter is measuring the same wave field as the buoy. Clearly the former records a spatial average over an instant whilst the latter makes a temporal average for a fixed location. However, in general, the altimeter does not pass right over the buoy, and when close to land the environmental conditions at the locations of the buoy and altimeter measurements may differ. In a detailed evaluation using output from a wave model, Nencioli and Quartly [30] showed that the "co-responsive" region around some coastal buoys could be very small, whereas for others it extended more than 50 km.

Figure 3 shows the geometry for a selection of the automated match-ups found within 50 km of land. Figure 3e shows an example of a good match-up between buoy location and altimeter pass, with land sufficiently far away to not affect either measurement. Figure 3a,f,g show cases where the nearest altimeter data to the buoy is far enough from land to be unaffected by it, but the buoys are much closer to the coast, so may register significantly different conditions, if the waves are affected by shoaling or if the prevailing wind-waves are away from the land. Figure 3b is an example of where a headland separates the region that was observed by the buoy from that sampled by the altimeter, and Figure 3c,d,h illustrate extreme cases where the altimeter measurements are from much more sheltered locations than the buoys and also the altimeter waveforms are likely to include extraneous reflections from land.

Figure 4 shows altimeter data along the first four of these sections. In the first panel, there are no spurious values in the altimeter section, as the locations are far enough from the influence of land and the median value agrees well with the buoy, despite being much closer to the coast and in shallower

conditions. However, in the other cases many large but valid measurements of SWH are returned by the altimeter, although there is much variability along-track. In Figure 4d, it is possible that the MLE-4 values around index 64210 are representative of true low SWH conditions, whilst the buoy, despite not being far off, is in much rougher conditions. The challenge is which of the cases should give reliable results and is there an objective way of selecting those buoys, without repeating the intricate modelling analysis of Nencioli and Quartly [30] separately for each buoy?
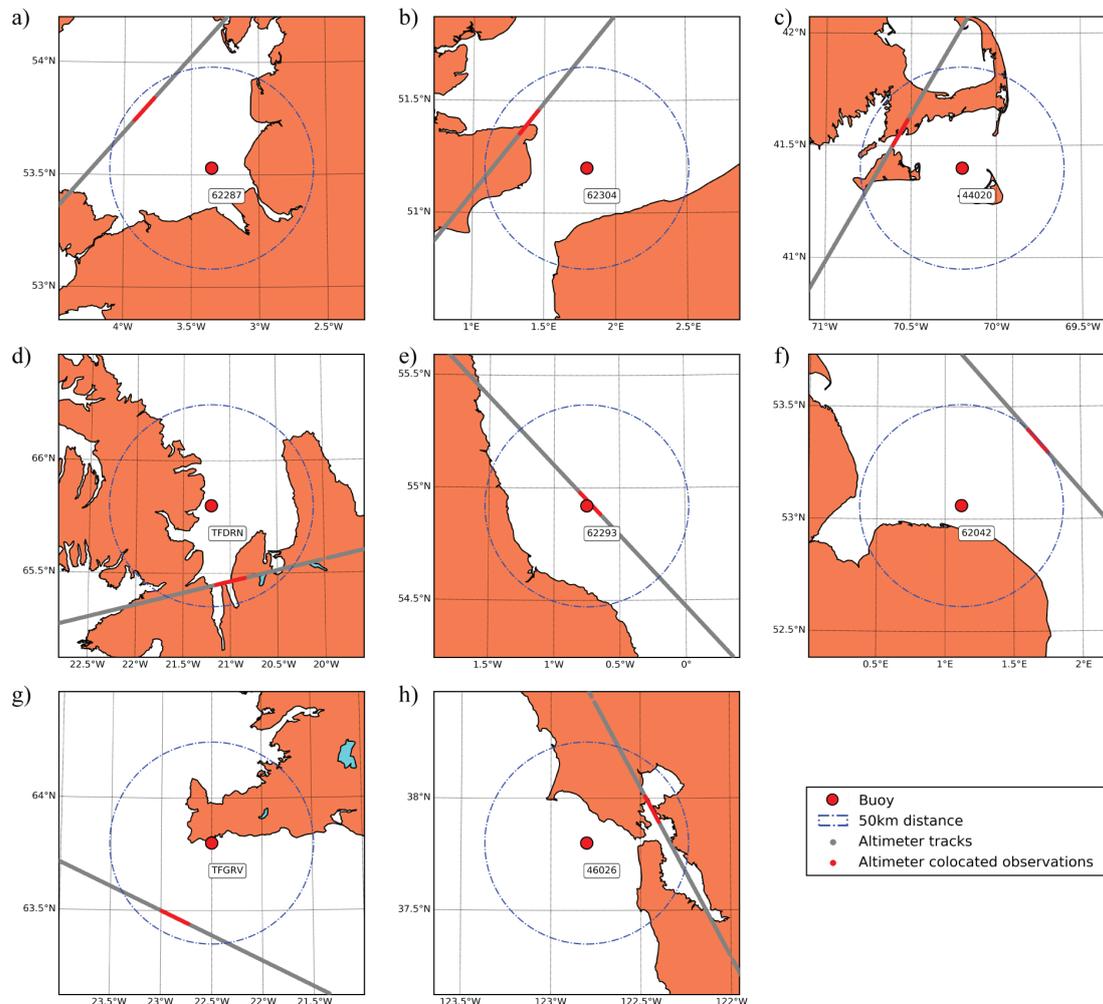
Figure 3. Locations of eight wave buoys, with nearest altimeter track by Jason-3, showing the variation in the quality of match-ups. (**a**) Buoy 62287 off Liverpool, UK. (**b**) Buoy 62304 in English Channel (**c**) Buoy 44020 off Barnstable, MA, USA. (**d**) Buoy TFDRN off NW Iceland. (**e**) Buoy 62293 off Humberside, UK. (**f**) Buoy 62042 off Norfolk, UK. (**g**) Buoy TFGRV off SW Iceland (**h**) Buoy 46026 off San Francisco, CA, USA. The blue circle indicates 50 km radius from the buoy, with the red segment of track indicating the 51 nearest points to the buoy.

Given that there is a strong societal need to use altimeter data in the coastal zone [31], then the algorithms need to be assessed in that environment. The initial Round Robin exercise described by Schlembach et al. [5] was mandated to use all the provided match-ups, whether or not the buoy and satellite measured the same wave-field. This meant that the error estimates for all algorithms were around ∼0.4 m, which is larger than normally noted [6,12], and it leaves the concern that the procedure may have affected some algorithms disproportionately. Thus, here we reassess the algorithms using a number of techniques developed with hindsight to develop a more realistic and robust guide as to how such analyses should be carried out. There are two main aspects we focus on: the minimum

number of high-frequency (20 Hz) satellite estimates needed to ensure useful altimeter values and the methodology for the selection of appropriate match-ups.
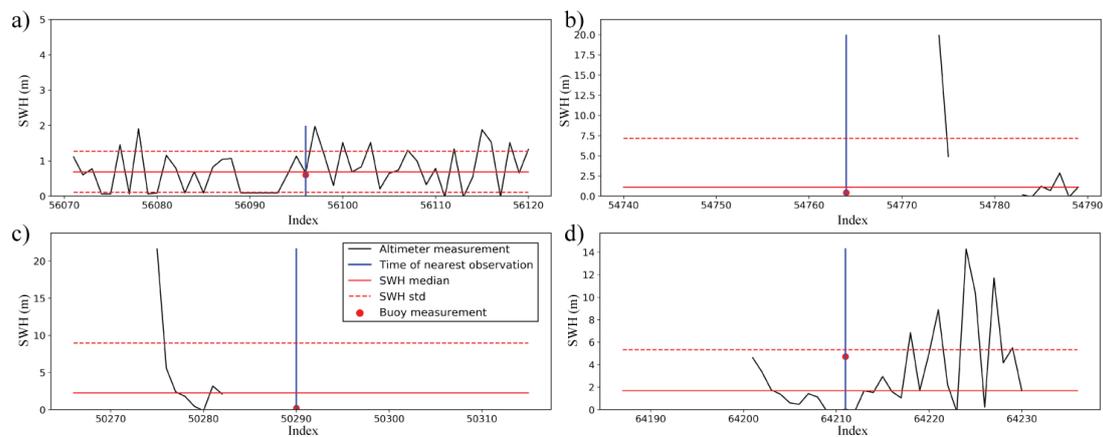
Figure 4. Examples of SWH profiles from MLE-4 for the nearest 51 points to each of four buoys; "index" refers to the records no. in data file. Points shown are those deemed "valid", with red lines showing the median and ±1 std. dev. of those estimates. The blue line marks the point of nearest approach to the buoy, with the red dot showing the buoy measurement. Locations are as in Figure 3a–d, namely (**a**) Buoy 62287 off Liverpool, UK. (**b**) Buoy 62304 in English Channel. (**c**) Buoy 44020 off Barnstable, MA, USA. (**d**) Buoy TFDRN in bay off NW Iceland.

3.1. Effect of Changing Minimum No. of Match Ups

The original documented procedure [11] had specified that after editing, the median of the nearest 51 20-Hz estimates to the buoy location would be used, without specifying a minimum number of valid records. The prescribed editing methodology was to discard observations over land, or flagged by the data provider, and to remove values outside the range −0.25 to 25 m or more than 3 std. dev. from a local 21-point running mean; the latter removes anomalies across intense rain cells [32] and ocean slicks.

Therefore, we repeated the buoy assessment of Schlembach et al. [5], but, in turn, requiring a minimum number of 1, 5, 10, or 20 valid estimates within the 51 observations, with the key characteristics (slope and intercept of best fit line, RMSE (root mean square error) and r^2) being noted. Figure 5 shows the results for the 11 different retracker algorithms applied to Jason-3 data. The top plots show the proportion of passes returning valid median values (as a %). In the coastal zone, none of the algorithms provide more than 42% of values for match ups, because some of the pairings corresponded to altimeter sections that were actually completely over land or the altimeter failed to record waveforms on that pass. The proportion of altimeter values drops to ~35% when the requirement is for at least 20 records to contribute. There are significant differences in the availability of an SWH estimate for the various algorithms, which is principally due to the data providers' recommended flagging: BrownP provides the greatest number of values, whilst STARv2, adapHF, and adapti provide the least.

More critical is examination of the r^2 and RMSE values, where WHALES, adapHF and adapti are the best for the coastal environment (Figure 5c,e), with STARv2, TALES, adapHF and adapti for the open ocean (Figure 5d,f). However, note that adapHF and adapti are among the worst when median is calculated from a minimum of one point, but among the best when five or more 20 Hz estimates are present. Figure 6 shows the progression for each algorithm as the minimum number of observations is raised. For the open ocean (Figure 6b), all algorithms improve in both measures (lowering RMSE and increasing r^2). For the coastal zone (Figure 6a), the situation appears more complicated, with RMSE always decreasing but the r^2 not rising uniformly; this is likely to be due to the exclusion of rare high SWH cases (as seen by both buoy and altimeter).

The algorithms that include the high-frequency adjustment to remove covariant errors have lower noise levels i.e., less short-term variability (not shown), but their representative average, whether from a minimum of 20 points or fewer has roughly the same accuracy (with respect to the buoys) as their parent product. The fact that adapHF performs slightly better than adapti is likely due to its correction term being based on a much wider spatial average than the 51 points used in calculating the altimeter measure. Indeed some other papers have calculated altimeter averages over much larger lengths than the 15–17 km advocated for this Round Robin exercise. As the STARv2 algorithm selects a very smooth profile through its multitude of solutions, the median value it deliver for such a section effectively takes input from a much wider region.

A similar analysis was carried out for SWH data from Sentinel-3A (Figures 7 and 8), with the results separated into retrackers which use the PLRM data and those that use the DDA waveforms. Some comparison with Jason-3 may be admitted, but it is important to recall that these analyses are for a different set of buoys, and a slightly different time period. In particular the documented pairings for S3A with buoys did not include sections predominantly over land as had been the case for Jason-3 pairings (see Figure 3), so the percentages of valid comparisons are much higher. In all cases, increasing the minimum number of valid points for the calculation of the median increases r^2 and reduces RMSE. A slight exception is SAMOSA (the default algorithm for S3A DDA waveform retracking) as so few points in the coastal zone are flagged as invalid (see Figure 2) that altimeter medians calculated using a minimum of 20 points are the same as those dependent on a minimum of one.
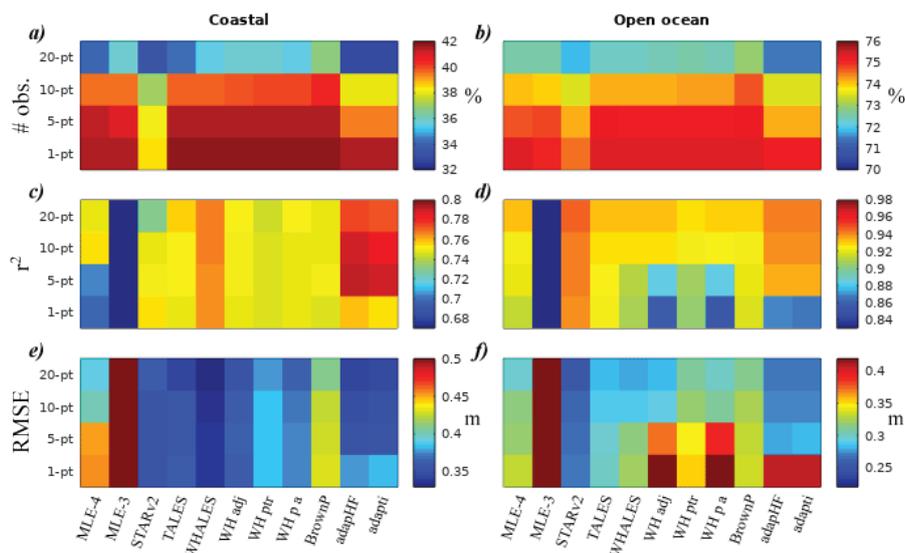
Figure 5. Effect of raising the minimum no. of observations on the number and quality of the median estimate by the Jason-3 altimeter. (**a**,**b**) show the percentage of passes exceeding the eligibility threshold; (**c**,**d**) show the r^2 values and (**e**,**f**) the RMSE upon comparison with buoys. Note that the bottom row is for a *minimum* of one point contributing to the median, and thus includes all the cases where 20 points are present plus those when between 1 and 19 are valid. Different colour scales are used for coastal (buoy within 15 km of coastline) and open ocean conditions to allow the variations in each panel to be clearly discerned.
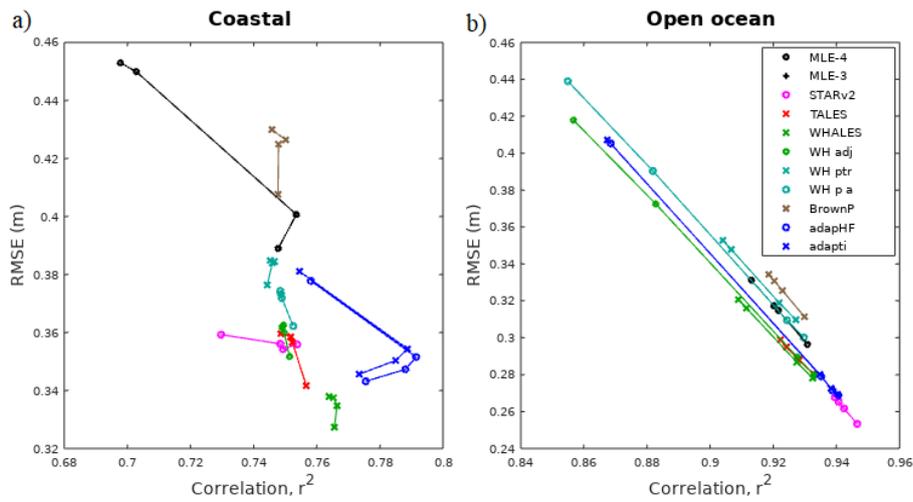
Figure 6. Generally improved performance as minimum no. of valid measurements is increased from 1 (at top left with low r^2 and high RMSE) to 5 to 10 to 20. Results are displayed for the various different Jason-3 algorithms for (**a**) Coastal zone (buoy within 15 km of coastline), (**b**) Open ocean. In most cases, more observations leads to higher r^2 and lower RMSE. (The chosen areas for these plots do not show MLE-3, which had much lower r^2 values and higher RMSE than the other algorithms.)
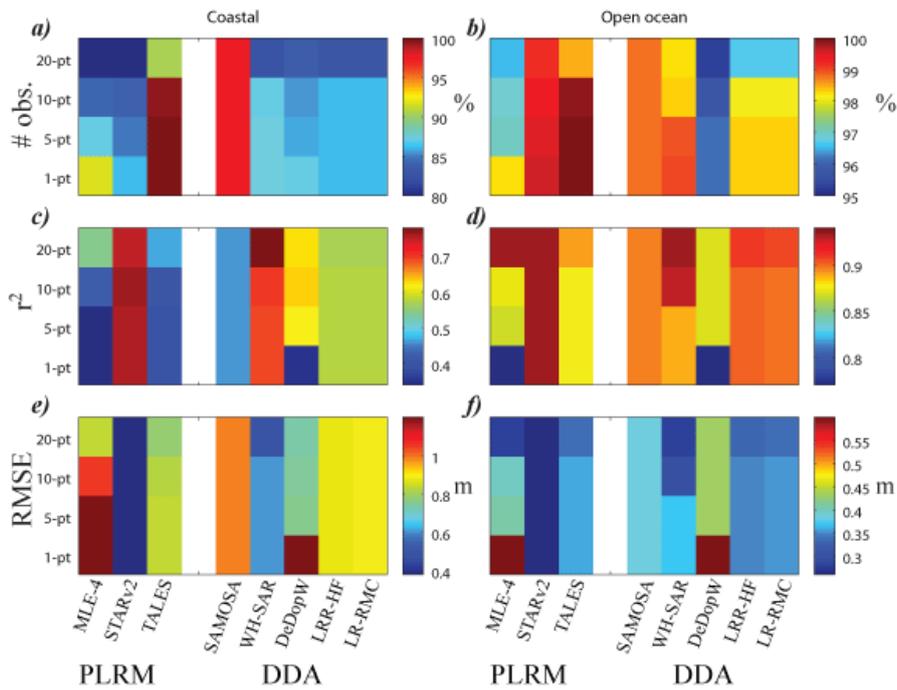
Figure 7. Effect of raising the minimum no. of observations on the number and quality of the median estimate by the Sentinel-3 altimeter. (**a,b**) show the percentage of passes exceeding the eligibility threshold; (**c,d**) show the r^2 values and (**e,f**) the root mean square error (RMSE) upon comparison with buoys. Note different colour scales are used for coastal (buoy within 15 km of coastline) and open ocean conditions to allow the variations in each panel to be clearly discerned.
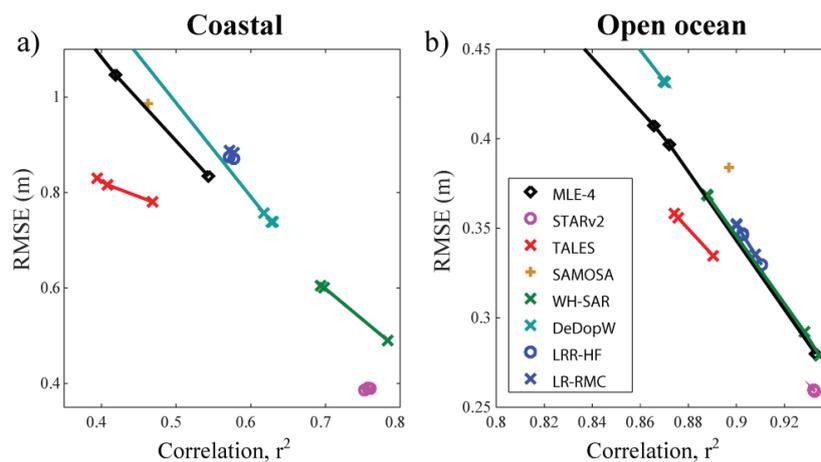
Figure 8. Generally improved performance as minimum no. of valid measurements is increased from 1 to 5 to 10 to 20. The results are displayed for the various different Sentinel-3A algorithms for (**a**) Coastal zone (buoy within 15 km of coastline), (**b**) Open ocean. For many algorithms, their performance with a minimum of one point led to such RMSE as to be plotted off the axes; in nearly all cases, more observations leads to higher r^2 and lower RMSE. Due to the limited flagging on SAMOSA, all its points coincide.

All of the algorithms give a better performance in the open ocean, despite the higher wave height conditions encountered, with STARv2 and WH-SAR being the best (Figure 8b). Again, we note that the way STARv2 selects a smooth SWH profile through its set of solutions implies that it will have already taken heed of information from a wider spatial extent. DeDopW and MLE-4 both perform particularly poorly if medians may be calculated from just a few points; increasing the threshold to five points removes the worst match ups, improving both r^2 and RMSE values. A similar situation occurs in the coastal zone, where STARv2 and WH-SAR are the standout performers (Figure 8a), and again DeDopW fares poorly if all representative averages are used, but is then the algorithm with the third lowest RMSE if medians are only determined from five or more valid estimates. From Figure 8a, one can see that with a minimum of 20 points contributing to the median WHALES-SAR delivers the highest r^2 for the coastal zone, but STARv2 gives the lowest RMSE. This plot also shows the relative improvement of DeDopW from outside the axes to third best algorithm for those conditions.

3.2. Automatic Selection of Best Buoys

The illustration of distant pairings of J3 and buoy data in Figure 3 show that, irrespective of land contamination of the altimeter signal, there may be great disparities in measurements due to the buoy and altimeter observations being in different wave environments. Whether the two estimates are equivalent depends upon the shape of the coastline, shoaling bathymetry, and the prevailing direction of waves. Furthermore, differently designed buoys may give different responses to a given wave field. We develop a method to identify a consistent set of buoy-altimeter pairings for comparison of algorithms so that our estimates of retracker error in the coastal zone are not dominated by these disparities.

The supplied dataset for Jason-3 had 7128 match-ups with individual passes, utilising 135 different buoy-ground track pairings. The two-year period being analysed corresponds to 73 repeats along each ground track, but there will be rare occasions when the altimeter is not operating nominally, and there will also be down-time for the buoys when they are serviced. Thus, we set a threshold that there should be a minimum of 40 coincident buoy-altimeter observations for a buoy to be considered. This removed some of those in the Arctic that only operated for a few months a year, plus some other intermittent instruments to leave us with 107, with many close to the coast (Figure 9).
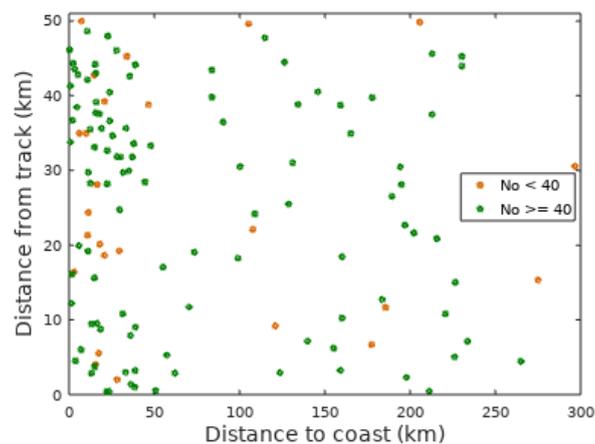
Figure 9. Contrasting measurements of distance from coast and distance from altimeter track. We only utilise those with at least 40 out of possible 73 Jason-3 repeats.

Although some of these buoys were further from the altimeter track than they were from the coast, this does not necessarily imply that they can not be used (see Figure 4a, for a counter-example). We proceed to select the most suitable buoys by comparison with the altimeter data. Clearly, if one selects buoys that perfectly match with your chosen altimeter algorithm, then the evaluation will be very favourable. Accordingly, instead, we take the median of the derived values for all the different algorithms, and use that to assess the consistency of the buoys. Subsequenty, these selected buoys can be used to assess and compare the different retrackers.

Thus, for each buoy-altimeter match up having at least 40 repeats, we perform a correlation analysis of the median of the different algorithms against the buoy measurements (the independent variable), determining the slope and bias of the best fit line and the r^2 value and the standard deviation (S.D.) of variations about this line (Figure 10). The panels have a bi-linear scale to allow greater focus on those buoys within 50 km of the coast. A number of those very close to the coast and far from the altimeter track have increased S.D. and reduced r^2, but by no means all. A co-representation of these derived metrics shows a consistency of the majority of buoys, but with a few outliers (Figure 11). The dashed lines show thresholds for acceptance of the match-up for the final cohort. The criteria have to be more lax for the coastal selection, or else there would be very few matchups used in that group. Interestingly, both boxes in Figure 11a are centred around a median slope of 0.925 suggesting that the majority of the retracker algorithms show less variation than the buoy data. These selection criteria, aiming to get a set of consistent buoys, reduced the number of open ocean pairings from 84 to 41 and the coastal number from 23 to 6, with the locations illustrated in Figure 12. (Note this does not necessarily imply that all the other buoys are "bad" or have a poor calibration; it is simply that they do not show an adequately good connection to the altimeter data, which may be due to the different exposure of the buoy and specified altimeter track, see Figure 3.)
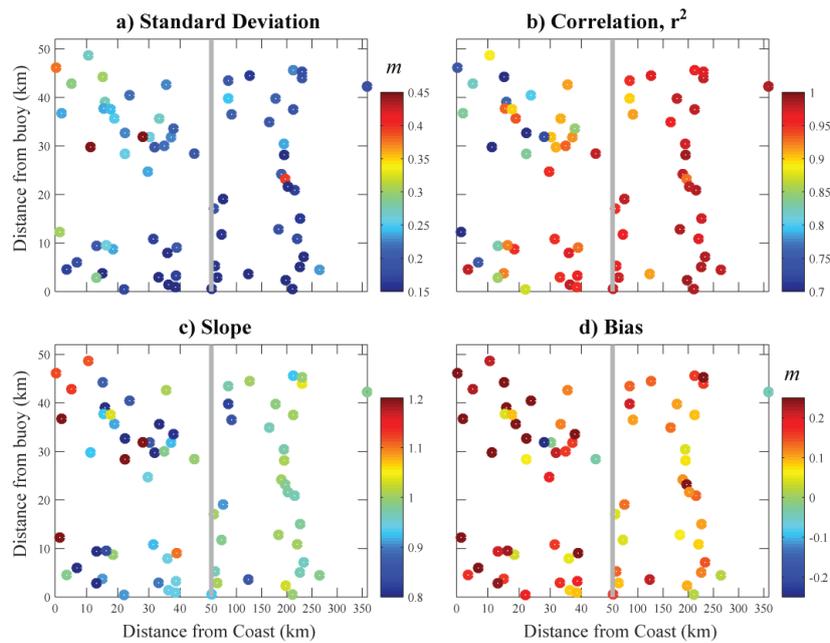
Figure 10. Assessment using the median of the estimates from the Jason-3 retrackers, showing the worse performance in general closer to the coast. (**a**) Standard deviation (m); (**b**) Correlation, r^2; (**c**) Slope; and, (**d**) Bias. (All plots have a bilinear x-axis to focus on the behaviour for those within 50 km of the coast).
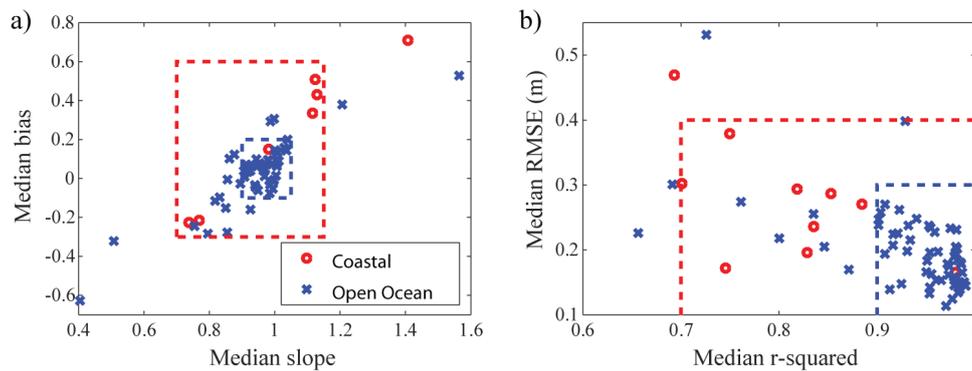
Figure 11. General assessment using the median of the estimates from the Jason-3 retrackers. (**a**) bias and slope values; (**b**) Correlation (r^2) and RMSE. In both plots, red points are for those within 15 km of the coast, and blue those further away, and the dashed lines show the selection criteria to discard outliers and give a consistent dataset.
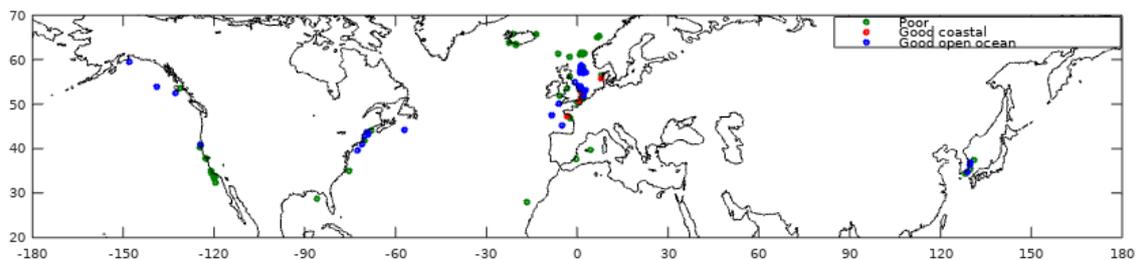
Figure 12. Locations of buoy-altimeter matchups. Red shows locations with good correlation in coastal region (less than 15 km from coast); blue for open ocean. Green marks those with poor correlation.

3.3. Effect on Assessment of Retrackers

In order to perform a robust evaluation of the SWH retrievals from the many retracking algorithms, we regress the values for each algorithm against the collated buoy data, either using all the buoy data available or only using the subset of match-ups in which we have the greatest confidence. The derived correlation coefficient (r^2) and RMSE for these different assessments are listed in Table 2, separated according to whether the buoys are within 15 km of the coast or further out.

The improvements in both r^2 and RMSE are quite marked for the open ocean data, with the error in SWH for MLE-4 decreasing from nearly 0.30 m to 0.20 m upon discarding of the least well-matched buoys, with similar fractional changes for the other retrackers. This was matched by changes in r^2 from ∼0.93 to ∼0.97. However, in the coastal regime, the chosen subset of buoys did increase r^2 marginally for all retrackers, but with the resultant RMSE also increasing slightly. For the coastal zone, the main improvement had been achieved by requiring at least 20 valid measurements; however, the differences due to the buoy and altimeter sampling different wave conditions remain. Almost all of the newly-developed retrackers do provide an improvement on MLE-4 in this regard; the exception is BrownP, which may be due to its flagging scheme rarely discarding any data as invalid (see Figure 2a).

Figure 13 illustrates these changes in performance. Mostly the data lie roughly along a diagonal line, showing that there is little change in the relative ranking of algorithms; however in the open ocean conditions, the STARv2 algorithm gave the best results (highest r^2 and lowest RMSE) when all the data were used, but after the buoy selection (which improved the results for all), it is now bettered by adapti and adapHF. For the most part, other algorithms retain their relative ranking.

A similar analysis was carried out for S3A. The database of match-ups contained 191 buoys, of which 96 were overflown 16 or more times Applying consistency checks, broadly similar to those for Jason-3 (see Figure 11) reduced the pairings in the coastal zone from 14 to 5, and those in the open ocean from 82 to 63. The results of the correlation analysis for the S3A retrackers with all the buoys and just the selected best buoys is given in Table 3.

All of the algorithms show a significant improvement (higher r^2, lower RMSE) upon using the dataset of selected buoys. This just highlights that there were a lot of poor pairings in the initial S3A match-up dataset, whether due to land contamination of the altimetry or the buoy residing in a distinct wave environment from that sampled by the altimeter. In the open ocean, the best retrackers are TALES and WH-SAR. Indeed, in general, the PLRM retrackers slightly outperform the DDA ones. There have been suggestions that a delay-Doppler altimeter is sensitive to other wave properties than just SWH [27], so this may explain the difference. For the coastal zone, the majority of the retrackers achieve an RMSE less than 0.25 m, which is impressive, but these results emanate from only five buoys, three or which were in the relatively low SWH conditions near the Strait of Dover. When all of the coastal buoys were used, STARv2 and WH-SAR were the standout retrackers. However. the process of refining the selection to a consistent set of buoys showed all to have very similar performance, with LRR-HF now having the highest r^2 and the second lowest RMSE.

Table 2. Results for the different retrackers applied to Jason-3 data showing the effect of different buoy selection criteria. In all cases, a minimum of 20 valid SWH measurements along track is required.

		r^2		
Retracker	All Coastal	Best Coastal	All Open Ocean	Best Open Ocean
MLE-4	0.748	0.770	0.931	0.971
MLE-3	0.547	0.583	0.790	0.859
STARv2	0.730	0.746	0.947	0.977
TALES	0.757	0.774	0.933	0.974
WHALES	0.766	0.787	0.933	0.975
WH adj	0.751	0.774	0.933	0.975
WH ptr	0.744	0.771	0.927	0.970
WH p a	0.753	0.772	0.930	0.972
BrownP	0.748	0.773	0.930	0.971
adapHF	0.775	0.810	0.941	0.980
adapti	0.773	0.809	0.940	0.980

		RMSE (m)		
Retracker	All Coastal	Best Coastal	All Open Ocean	Best Open Ocean
MLE-4	0.389	0.400	0.296	0.204
MLE-3	0.564	0.567	0.547	0.471
STARv2	0.359	0.378	0.253	0.178
TALES	0.342	0.359	0.280	0.189
WHALES	0.327	0.336	0.278	0.182
WH adj	0.352	0.360	0.280	0.185
WH ptr	0.376	0.382	0.310	0.213
WH p a	0.362	0.374	0.300	0.202
BrownP	0.408	0.416	0.311	0.213
adapHF	0.343	0.342	0.269	0.167
adapti	0.346	0.344	0.269	0.168
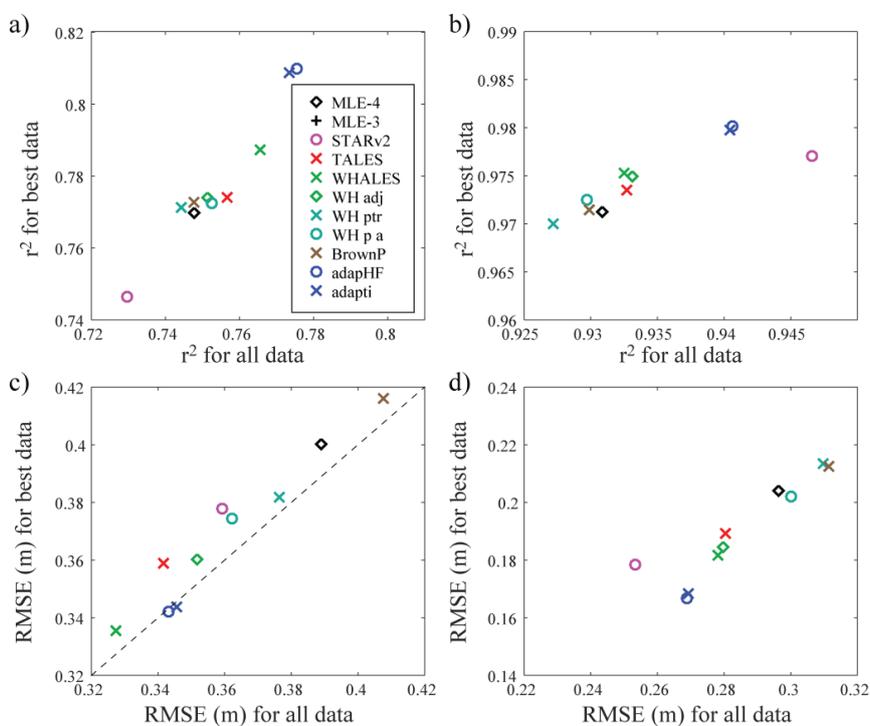
Figure 13. Effect of data selection on the metrics r^2 and RMSE. (**a,c**) Coastal; (**b,d**) Open ocean. MLE-4 is the default algorithm for Jason-3, and MLE-3 retrievals are much worse than the others (lower r^2 and higher RMSE), so they are not in the domains focussed on in each panel. Note those algorithms indicated by a circle have some implicit along-track smoothing in them.

It is tempting to compare the results for Jason-3 and Sentinel-3A e.g., noting that the MLE-4, STARv2, and TALES algorithms applied to the best buoys for S3A give better results in the coastal zone (higher r^2, lower RMSE) than those algorithms applied to Jason-3, whereas the converse is true for the best open ocean buoys. However, it should be remembered that different selections of buoys have been used in each case, with different proximity to the tracks and slightly different periods of coverage and, thus, the range of conditions experienced.

Table 3. The results for the different retrackers applied to Sentinel-3A data showing the effect of different buoy selection criteria. In all cases, a minimum of 20 valid SWH measurements along track is required. The SAMOSA retracker is omitted here because of uncertainty about the implementation of its quality flag.

| | r^2 | | | |
Retracker	All Coastal	Best Coastal	All Open Ocean	Best Open Ocean
MLE-4	0.612	0.902	0.934	0.961
STARv2	0.752	0.896	0.932	0.957
TALES	0.468	0.904	0.890	0.962
WH-SAR	0.783	0.915	0.935	0.966
DeDopW	0.627	0.901	0.870	0.963
LR-HF	0.571	0.925	0.911	0.953
LR-RMC	0.571	0.923	0.908	0.951
	RMSE (m)			
Retracker	All Coastal	Best Coastal	All Open Ocean	Best Open Ocean
MLE-4	0.768	0.274	0.276	0.215
STARv2	0.386	0.232	0.259	0.210
TALES	0.780	0.241	0.335	0.198
WH-SAR	0.490	0.239	0.279	0.203
DeDopW	0.738	0.276	0.432	0.231
LR-HF	0.875	0.238	0.330	0.241
LR-RMC	0.888	0.243	0.335	0.249

4. Summary and Discussion

There have been many papers advocating new algorithms for estimating SWH from altimeters, with buoy comparisons usually including MLE-4 as a reference, as has been a standard for many years and is disseminated with most altimeter datasets. However, an intercomparison of the results stated in these papers is challenging because of how much the chosen metrics depend upon the particular selection of buoys and the flagging of the altimeter data. The Round Robin exercise detailed by Schlembach et al. [5] provided a major step forward in the relative assessment of algorithms, but some aspects were sub-optimal, because, for very good reasons, the details of the procedure were fully specified ahead of any examination of the data.

In particular, it was noted that the comparisons using buoys in the coastal zone showed RMSE values of >0.6 m for most J3 retrackers, which were potentially linked to the insistence on using all altimeter estimates not flagged as invalid, and the inclusion of buoys with poor calibration or with different exposure to the wave field. This paper has explored these issues.

As the locations of the altimeter data to be averaged were identified through an automatic match-up process, those points nearest to a coastal buoy often included a number of records over or very close to land. Although the data editing discarded all points over land, the flagging of suspect data close to land was left to the individual developers, with some electing for a cautious approach whilst others prioritised a high return rate of acceptable data. We used the median as a robust measure of the average, but, in some circumstances, over half the valid points may be affected (e.g., Figure 4c). This problem could be reduced by insisting that a higher number of observations

contributed to the median (Figure 5). Such an alteration of the protocol improved the assessment of all the Jason-3 retracker algorithms, although the improvements were not as large as some of the differences between algorithms. For the open ocean, raising the threshold on the number of points increased r^2 and reduced RMSE for every algorithm; in the coastal zone, the changes in r^2 were not monotonic (Figure 6a). Similar analysis for Sentinel-3A data (Figures 7 and 8) showed increasing the minimum led to improvements in assessment for all the new algorithms, with SAMOSA (the standard product) showing no change, because it had so few data marked as invalid that no cases were removed by raising the threshold.

As expected, the choice of open ocean environment or coastal zone affects the relative ranking of algorithms, because some retrackers had been particularly designed with the coastal regime in mind. More surprisingly, the threshold on the minimum number of 20 Hz estimates to be averaged also had a significant impact on which algorithms were perceived as best. This was because, with limited data flagging, a median calculated from only a few suspect estimates could still be markedly wrong, whereas a requirement for at least 20 valid values present tended to include a higher proportion of accurate estimates.

That analysis was effectively a selection of measurements for each individual pass along the track. Section 3.3 explored which buoy-track pairings contributed most to the observed errors. A pairing selection procedure was devised using the median of the different retracker estimates to elucidate which were the good match-ups without favouring a particular retracker. Figure 10a,b show that the RMSE and r^2 associated with a buoy more than 50 km from the coast did not vary appreciably with the separation between the buoy and altimeter observations. The majority of pairings closer to the coast than 30 km showed slightly reduced r^2 values and higher RMSE, especially when coupled with a separation between buoy and altimeter measurement exceeding 30 km. The buoy pairings giving the most extreme performance were then discarded.

The set of buoys from this selection process was then used to assess the various algorithms. For Jason-3, the biggest improvements were noted for the open ocean, with most retrackers now having an RMSE of less than 0.2 m (Table 2), whereas, somewhat surprisingly, the changes were minimal in the coastal zone. This is probably because the first step (requiring a minimum of twenty 20 Hz estimates) had already removed most of the poor pairings. However, despite all of these changes, the ranking of the algorithms remained similar, apart from the relative standing of STARv2 and adapti/adapHF for the open ocean. For Sentinel-3A, the focussing on the best buoys led to large improvements in the coastal zone as well as in the open ocean. Given that there is less established heritage for the Sentinel-3 data, this may indicate that those developers had yet to optimise their flagging. Certainly, it demonstrates the benefits to be gained by having requirements on the number of valid altimeter estimates *and* on the consistency of the buoy data.

Whilst it is easy to make an assessment of an altimeter retracker by a comparison with a set of buoys, it is hard to compare the values achieved by different researchers using different satellites, retrackers, and networks of buoys, as well as subtly different methodologies. That is why detailed multi-algorithm intercomparisons, such as advocated in the ESA Climate Change Initiative programme, are key to determining the best algorithms for generation of long-term datasets. Simple buoy-satellite comparisons furnish the magnitude of the discrepancies between them; to partition this into the error associated with the buoys and with the altimeter requires triple collocation analysis, necessitating a third totally independent contemporaneous dataset [33].

If all of the error sources are independent, the variance of the mismatch between buoys and retracked altimeter data should be the sum of the variances due to buoy measurement, retracker error, and space-time sampling differences and, thus, the inclusion or not of poorly calibrated or situated buoys would not affect the ranking. Although our results broadly confirm this behaviour, the selection of a best coastal retracker does depend, to some extent, upon the processing and buoy selection criteria. This is significantly affected by the proposed data flagging, as some algorithms provide an estimate in nearly all conditions, whilst others are more cautious (Figure 2). Insisting that sufficient observations

contribute to the altimeter average removes the contribution of the most egregious altimeter-buoy pairings and favours those algorithms that flag potentially poor estimates. The original Sea State CCI Round Robin attempted to follow a preordained protocol in order to avoid the perception that the procedure was biased towards any particular provider's product. However, a slight modification (discarding buoys that gave consistently large discrepancies, see Section 3.6 of [5]) was needed in order to give r^2 and RMSE values commensurate with other published studies. This highlights the pitfalls of fully characterising a procedure before an initial viewing of the data. In this paper, we have documented the issues, and provided a template for how others working on Round Robin investigations can excise data match-ups that affect the majority of algorithms being assessed.

Abbreviations

The following abbreviations are used in this manuscript:

ALES	Adaptive Leading-Edge Subwaveform
CCI	Climate Change Initiative
DDA	Delay-Doppler Altimetry
ESA	European Space Agency
J3	Jason-3
LR-RMC	Low-Resolution with Range Migration Correction
LRM	Low Resolution Mode
MLE	Maximum Likelihood Estimator
PLRM	Pseudo Low Resolution Mode
RMSE	Root mean square error
S3A	Sentinel-3A
SAR	Synthetic Aperture Radar
S.D.	Standard Deviation
STAR	Spatio-Temporal Altimetry Retracking
SWH	Significant Wave Height

References

1. Global Climate Observing System. Systematic Observation Requirements for Satellite-Based Data Products for Climate (2011 Update). Available online: https://library.wmo.int/doc_num.php?explnum_id=3710 (accessed on 1 May 2020).
2. Ardhuin, F.; Stopa, J.E.; Chapron, B.; Collard, F.; Husson, R.; Jensen, R.E.; Johannessen, J.; Mouche, A.; Passaro, M.; Quartly, G.D.; Swail, V.; Young, I. Observing sea states. *Front. Mar. Sci.* **2019**, *6*, 124. [CrossRef]
3. Young, I.R.; Ribal, A. Multiplatform evaluation of global trends in wind speed and wave height. *Science* **2019**, *364*, 548–552. [CrossRef] [PubMed]
4. Dodet, G.; Piolle, J.F.; Quilfen, Y.; Abdalla, S.; Accensi, M.; Ardhuin, F.; Ash, E.; Bidlot, J.R.; Gommenginger, C.; Marechal, G.; et al. The Sea State CCI dataset v1: Towards a sea state climate data record based on satellite observations. *Earth Syst. Sci. Data Discuss.* **2020**, *2020*, 1–28. [CrossRef]

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

6. Durrant, T.H.; Greenslade, D.J.M.; Simmonds, I. Validation of Jason-1 and Envisat remotely sensed wave heights. *J. Atmos. Ocean. Technol.* **2009**, *26*, 123–134. [CrossRef]

7. Zieger, S.; Vinoth, J.; Young, I.R. Joint calibration of multiplatform altimeter measurements of wind speed and wave height over the past 20 years. *J. Atmos. Ocean. Technol.* **2009**, *26*, 2549–2564. [CrossRef]

8. Yang, J.; Zhang, J. Validation of Sentinel-3A/3B satellite altimetry wave heights with buoy and Jason-3 data. *Sensors* **2019**, *19*, 2914. [CrossRef]

9. Gomez-Enri, J.; Vignudelli, S.; Quartly, G.D.; Gommenginger, C.P.; Cipollini, P.; Challenor, P.G.; Benveniste, J. Modeling Envisat RA-2 waveforms in the coastal zone: Case study of calm water contamination. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 474–478. [CrossRef]

10. Wang, X.; Ichikawa, K. Coastal waveform retracking for Jason-2 altimeter data based on along-track echograms around the Tsushima Islands in Japan. *Remote Sens.* **2017**, *9*, 762. [CrossRef]

11. ESA. Round Robin: Final Selection and Ranking of Algorithms. Available online: http://cci.esa.int/sites/default/files/Sea_State_cci_RR_Final_Selection_v1.1-signed.pdf (accessed on 3 July 2020).

12. Queffeulou, P. Long-term validation of wave height measurements from altimeters. *Mar. Geod.* **2004**, *27*, 495–510. [CrossRef]

13. Challenor, P.G.; Cotton, P.D. The Joint Calibration of Altimeter and In Situ Wave Heights. Advances in the Application of Marine Climatology. Available online: https://www.wmo.int/pages/prog/amp/mmop/documents/JCOMM-TR/J-TR-10-CLIMAR-99/Presentations/session_03/03.2_Challenor.pdf (accessed on 13 July 2020).

14. Raney, R.K. The delay/Doppler radar altimeter. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1578–1588. [CrossRef]

15. Quartly, G.D.; Nencioli, F.; Raynal, M.; Bonnefond, P.; Nilo Garcia, P.; Garcia-Mondéjar, A.; Flores de la Cruz, A.; Crétaux, J.F.; Taburet, N.; Frery, M.L.; et al. The roles of the S3MPC: Monitoring, validation and evolution of Sentinel-3 altimetry observations. *Remote Sens.* **2020**, *12*, 1763. [CrossRef]

16. Amarouche, L.; Thibaut, P.; Z.Zanife, O.; Dumont, J.P.; Vincent, P.; Steunou, N. Improving the Jason-1 ground retracking to better account for attitude effects. *Mar. Geod.* **2004**, *27*, 171–197. [CrossRef]

17. Passaro, M.; Cipollini, P.; Vignudelli, S.; Quartly, G.D.; Snaith, H.M. ALES: A multi-mission adaptive subwaveform retracker for coastal and open ocean altimetry. *Remote Sens. Environ.* **2014**, *145*, 173–189. [CrossRef]

18. ESA. Algorithm Theoretical Basis Document. Available online: http://cci.esa.int/sites/default/files/Sea_State_cci_ATBD_v1.1-signed_0.pdf (accessed on 28 June 2020).

19. Dinardo, S.; Fenoglio-Marc, L.; Buchhaupt, C.; Becker, M.; Scharroo, R.; Fernandes, M.J.; Benveniste, J. Coastal SAR and PLRM altimetry in German Bight and West Baltic Sea. *Adv. Space Res.* **2018**, *62*, 1371–1404. [CrossRef]

20. Peng, F.; Deng, X. A new retracking technique for Brown Peaky altimetric waveforms. *Mar. Geod.* **2018**, *41*, 99–125. [CrossRef]

21. Peng, F.; Deng, X. Validation of improved significant wave heights from the Brown-Peaky (BP) retracker along the east coast of Australia. *Remote Sens.* **2018**, *10*, 1072. [CrossRef]

22. Poisson, J.; Quartly, G.D.; Kurekin, A.A.; Thibaut, P.; Hoang, D.; Nencioli, F. Development of an ENVISAT altimetry processor providing sea level continuity between open ocean and Arctic leads. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5299–5319. [CrossRef]

23. Sandwell, D.T.; Smith, W.H.F. Retracking ERS-1 altimeter waveforms for optimal gravity field recovery. *Geophys. J. Int.* **2005**, *163*, 79–89. [CrossRef]

24. Quartly, G.D. Removal of covariant errors from altimetric wave height data. *Remote Sens.* **2019**, *11*, 2319. [CrossRef]

25. Roscher, R.; Uebbing, B.; Kusche, J. STAR: Spatio-temporal altimeter waveform retracking using sparse representation and conditional random fields. *Remote Sens. Environ.* **2017**, *201*, 148–164. [CrossRef]

26. Ray, C.; Martin-Puig, C.; Clarizia, M.P.; Ruffini, G.; Dinardo, S.; Gommenginger, C.; Benveniste, J. SAR altimeter backscattered waveform model. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 911–919. [CrossRef]

27. Moreau, T.; Tran, N.; Aublanc, J.; Tison, C.; Le Gac, S.; Boy, F. Impact of long ocean waves on wave height retrieval from SAR altimetry data. *Adv. Space Res.* **2018**, *62*, 1434–1444. [CrossRef]

28. Makhoul, E.; Roca, M.; Ray, C.; Escolà, R.; Garcia-Mondéjar, A. Evaluation of the precision of different delay-Doppler processor (DDP) algorithms using CryoSat-2 data over open ocean. *Adv. Space Res.* **2018**, *62*, 1464–1478. [CrossRef]

29. Monaldo, F. Expected differences between buoy and radar altimeter estimates of wind speed and significant wave height and their implications on buoy-altimeter comparisons. *J. Geophys. Res. Ocean.* **1988**, *93*, 2285–2302. [CrossRef]

30. Nencioli, F.; Quartly, G.D. Evaluation of Sentinel-3A wave height observations near the coast of southwest England. *Remote Sens.* **2019**, *11*, 2998. [CrossRef]

31. Passaro, M.; Fenoglio-Marc, L.; Cipollini, P. Validation of significant wave height from improved satellite altimetry in the German Bight. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2146–2156. [CrossRef]

32. Guymer, T.H.; Quartly, G.D.; Srokosz, M.A. The effects of rain on ERS-1 radar altimeter data. *J. Atmos. Ocean. Technol.* **1995**, *12*, 1229–1247.<1229:TEOROR>2.0.CO;2. [CrossRef]

33. Janssen, P.A.E.; Abdalla, S.; Hersbach, H.; Bidlot, J.-R. Error estimation of buoy, satellite, and model wave height data. *J. Atmos. Ocean. Technol.* **2007**, *24*, 1665–1677. [CrossRef]