IEEE Transactions on Geoscience and Remote Sensing

Development of an ENVISAT altimetry processor providing sea level continuity between open ocean and Arctic leads

Journal:	*Transactions on Geoscience and Remote Sensing*
Manuscript ID	TGRS-2017-00491.R2
Manuscript Type:	Regular paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Poisson, Jean-Christophe; Collecte Localisation Satellite (CLS), Direction Oceanographie Spatiale (DOS) Quartly, Graham; Plymouth Marine Laboratory, Earth Observation Kurekin, Andrey; Plymouth Marine Laboratory, Earth Observation Thibaut, Pierre; Collecte Localisation satellite, Space Oceanography Division Hoang, Duc; Collecte Localisation Satellite (CLS), Direction Oceanographie Spatiale (DOS) Nencioli, Francesco; Plymouth Marine Laboratory, Marine Earth Observation
Keywords:	Remote sensing, Radar altimetry, Arctic regions, Algorithms, Sea surface, Sea ice

Development of an ENVISAT altimetry processor providing sea level continuity between open ocean and Arctic leads

J.-C. Poisson[1], G. Quartly[2], A. Kurekin[2], P. Thibaut[1], D. Hoang[1], F. Nencioli[2]
[1]*CLS,* [2]*PML*

December 20, 2017

Abstract

Over the Arctic regions, current conventional altimetry products suffer from a lack of coverage or from degraded performance due to the inadequacy of the standard processing applied in the ground segments. This paper presents a set of dedicated algorithms able to process consistently returns from open ocean and from sea ice leads in the Arctic Ocean (detection of water surfaces and derivation of water levels using returns from these surfaces). This processing extends the area over which a precise sea level can be computed. In the frame of the ESA Sea Level Climate Change Initiative (CCI, http://cci.esa.int), we have first developed a new surface identification method combining two complementary solutions, one using a multiple criteria approach (in particular the backscattering coefficient and the peakiness coefficient of the waveforms) and one based on a supervised neural network approach. Then, a new physical model has been developed (modified from the Brown model to include anisotropy in the scattering from calm protected water surfaces) and has been implemented in a Maximum Likelihood Estimation retracker. This allows us to process both sea-ice lead waveforms (characterized by their peaky shapes) and ocean waveforms (more diffuse returns), guaranteeing, by construction, continuity between open ocean and ice-covered regions. This new processing has been used to produce maps of Arctic sea level anomaly from 18Hz ENVISAT/RA-2 data.

1 INTRODUCTION

The Arctic is an important component of the climate system whose exact influence on the global oceanic and atmospheric circulation is still not well known. It is also a very sensitive region to global warming and some of its direct effects such as ice melting are already particularly visible (Steele et al. [2008], Kwok and Rothrock [2009], Morison et al. [2012]). In this context, knowledge of the variability of a field such as the sea level in the Arctic Ocean and of the mechanisms which are responsible for it would enable us to better understand the rapid changes at work in this region. For more than 20 years, satellite altimetry has been recognized as the most accurate technique to measure sea surface height (SSH) at scales ranging from basins down to the mesoscale regimes (Le Traon [2013]), with the gradients in SSH providing quantitative values for the surface geostrophic currents (Fu and Cazenave [2000]). As large parts of the Arctic Ocean are still regularly covered by sea-ice, there are few sources of in situ data that can contribute to monitoring such a climatically important environment. The mean ice extensions during March (sea-ice maximum) and September (end of the ice-melting period) are shown in Fig. 1. This reveals that only the North Atlantic is totally ice-free. Coastal areas where freshwater fluxes from river runoffs have a strong influence (Armitage et al. [2016]; Serreze et al. [2006]) have seasonal ice cover. The central Arctic region has been, until recently, permanently ice-covered including the Beaufort Gyre which is a major feature of the Arctic circulation (Giles et al. [2012]). Studying and understanding the dynamic circulation of the Arctic thus necessitates the development of a SSH retrieval system that operates consistently through the changes between open ocean and floes with leads: this paper constructs such a product using nearly 10 years of ENVISAT altimetry data.

Radar altimeters emit a rapid series of pulses and record the resultant reflections from the Earth's surface. A wind-roughened ocean surface will have a wide region of reflecting facets contributing to the overall return echo. This waveform from a diffuse set of reflectors has a broad shape (Fig. 2a), which is described by the Brown model (Brown [1977]). Geophysical informations are usually derived by fitting a simple mathematical form, using a processing called "retracking" relating geophysical variables to the parameters controlling the shape and position of the waveform (Brown [1977]; Hayne [1980]). Over the ocean the amplitude of the signal relates to the mean square slope of the sea surface, which is due to wind; the slope of the leading edge conveys information about the significant wave height; and the position of the leading edge (see Fig.2a) informs us about the distance to the sea surface from which we determine the sea surface height (SSH). For normal ocean processing, the retracked point (which provides the range) should be close to the position of half-power (corresponding to the median height of reflecting facets); in other contexts (e.g. ice sheet processing) a lower threshold may be used (Bamber [1994]).

The European Space Agency (ESA) has set up the Climate Change Initiative (CCI, http://cci.esa.int) to construct and maintain consistent long-term datasets of essential climate variables. The component focusing on sea level has already completed its first two phases (Ablain et al. [2015], Quartly et al. [2017]) showing great improvements in the quality of the dataset in the Arctic using the algorithm presented in this paper. Indeed, measurements are severely impacted by the presence of sea-ice affecting the results of standard altimetry processing. We must note that most of the previous studies using altimetry data in the Arctic Ocean have been devoted to sea ice characterization (sea ice extent, freeboard and ice thickness estimation, ...). Very few have been focused on sea level determination with the constraint to ensure continuity of the observations between deep ocean and sea-ice regions, which is based on the computation of geographical bias between sea level estimates in the two areas (Peacock and Laxon [2004], Giles et al. [2012], Armitage et al. [2016]). It is the objective of this paper to present the processing allowing to provide accurate sea level measurements in the whole Arctic Ocean. We can also note that this processing will be of great interest for the determination of ice thickness, based on the computation of the freeboard height, itself linked to the precise determination of sea level in the leads enclosed in sea ice.

Ice floes within the radar waveform footprint affect the accuracy of measurements derived from standard altimetry processing. Firstly this may be because radar waves backscattered by the top of the floes form an ocean-like echo (but at a range corresponding to the surface of the floes and not to the surface of the sea). Secondly the altimeter footprint may contain both floes and leads or polynyas (area of open water surrounded by sea ice), generating complex waveforms with specular reflection from the near-glassy surface within them producing a very different waveform (Fig. 2b). Some researchers e.g. Prandi et al. [2012] and Cheng et al. [2015], have investigated changes in Arctic sea level using only standard altimetry processing and a careful selection of the data. However, better coverage and more reliable data can be produced using a processing scheme that accounts for the different shapes of waveforms in the Arctic.

1.1 Previous approaches to altimetry in sea-ice regions

One of the first processing dedicated to sea ice has been developed by Laxon [1994] (and Laxon and McAdoo [1994]). It was based on the identification and the retracking of altimeter waveforms, which allowed the estimation of sea level and sea-ice thickness in the Arctic from ERS-1 data. The method, based on the analysis of altimeter measurements, distinguishes leads from ice floes depending on their radar echo shape (an illustration of the backscattered power over lead is given in Fig 2b while Fig. 2c gives the various possible conventional altimeter returns) and then estimates the sea surface height using an empirical retracker also called "threshold retracker". This processing has been further implemented by Peacock and Laxon [2004] using ERS-1 and ERS-2 measurements to determine the first altimeter-derived sea surface height variability map of the Arctic Ocean. In that paper, the authors explain that even though residual errors in sea surface height estimates are greater in ice-covered region

than for ice-free regions, they are small enough to carry out geophysical analyses from these data. An update of the lead retracking step has been developed by Giles et al. [2007] and was used in Laxon et al. [2013] with CRYOSAT-2 data. The simple leading-edge threshold algorithm has been replaced by an empirical Gaussian-plus-exponential model to fit the echo in order to reduce the estimation noise. For the retracking of the CRYOSAT-2 Delay/Doppler waveforms over leads, Kurtz et al. [2014] has recently proposed a physical model based on variation of the backscattering coefficient (σ^0) with incidence angle. The use of a physical approach to fit returns from leads allows estimates to be less sensitive to the combined effect of bandwidth-limited range resolution and surface roughness variations compared with empirical retrackers. A similar approach is developed in this paper but for conventional altimeter returns. It is detailed in the third section of this paper.

1.2 The Arctic Ocean datasets

In this paper, we develop an end-to-end processing system able to yield estimates of sea level in both ice-free and ice-covered Arctic Ocean using ENVISAT RA-2 data. The ENVISAT satellite was launched by ESA in March 2002, and has been operating until April 2012. It principally flew in a 35-day repeat orbit, with a high inclination (82°), giving it coverage of the majority of the Arctic Ocean. Its radar altimeter, termed RA-2, operated in Ku-band (13.6 GHz) and S-band (3.2 GHz), with the data at the former frequency being used here due to their greater precision (higher bandwidth). The RA-2 provided average waveforms at 18 Hz (≈370 m along track), each of them being obtained by summation of 100 independent pulses and sampled over 128 bins of 3.125 ns (with 2 additional bins at the beginning of the waveform leading edge called "DFT points"). For this work, we make use of the Sensor Geophysical Data Records (SGDR), which contain all the waveform information plus all the ancillary corrections required to compute an accurate Sea Surface Height (SSH). More details on the operation of radar altimeters can be found in Fu and Cazenave [2000] and specific informations on RA-2 are detailed in Benveniste et al. [2002].

The system is based on the general methodology already mentioned and defined by Peacock and Laxon [2004]. The processing scheme, summarized in Fig. 3, commences with an initial selection and editing (quality control) of the data. Two independent approaches for the waveform classification are used (detailed in sections 2.2 and 2.3), which are then combined into a hybrid approach, marrying the cautious nature of one with the greater data coverage of the other. A new mathematical model for tracking the waveforms associated with leads (Fig. 2b) is then introduced in section 3. This model (based on a modified version of the Brown model commonly used for ocean waveforms) allows us to process ice-free and lead measurements using the same unique retracker. A further data editing step is described in section 4, as the overall approach is designed to be very conservative, keeping only the best data rather than accepting a large quantity of possibly useful measurements. The remit of the CCI project is to produce monthly averages of sea level, rather than to map the mesoscale variability at scales smaller than 50 km. The final section is a summary of the methodology espoused in this paper; the results and oceanographic interpretation of the reprocessed CCI data will be published subsequently.

2 CLASSIFICATION OF REFLECTING SURFACES

2.1 Background on cause of different waveform shapes

The first key step for extending sea level estimation in the Arctic sea-ice region is to discriminate between measurements over water, where an altimeter range may be retrieved and converted into an estimate of sea level, and reflections from ice floes that must be ignored (but the same method would allow us to process sea ice reflections and to derive freeboard heights and ice thicknesses by differentiation with water level). For a conventional altimeter such as ENVISAT, the waveforms are built up as the sum of the backscattered power from the reflectors at the

Earth's surface in a series of concentric annuli, with the strength of the reflection from each part of the surface being directly linked to the roughness of the water surface. For a homogeneous slightly rough surface, the relative strength of the reflected signal from successive annuli is principally controlled by the instrument beamwidth, leading to the Brown waveform described in Fig. 2a. The assumption of surface backscattering homogeneity, a strong hypothesis for standard ocean models Brown [1977]; Hayne [1980] is generally true over the open ocean (in nominal conditions), but much less in sea ice regions or in coastal regimes for example where bays or polynya protected from the wind may produce a near-glassy surface and thus specular returns quite different than the Brown waveform Gomez-Enri et al. [2010].

In the Arctic environment, a uniform cover of sea-ice can produce Brown-like returns, although typically with a much stronger signal due to the greater reflectivity of sea-ice with respect to water. Indeed, Peacock and Laxon [2004] and Drinkwater [1991] noted that consolidated ice such as fast ice or vast floes generate an isotropic rough surface. But a conventional altimeter often observes at the same time a large variety of reflecting surfaces (Laxon [1994]). In particular leads and polynyas mainly composed of calm water or new sea ice can produce altimeter waveforms that look like an impulse function as shown in Fig. 2b. Between these two extrema, complex waveform shapes are observed over sea-ice regions with varying contributions from solid ice floes, new ice and protected open water. Strong rain events can also cause unusual waveforms shapes (Guymer et al. [1995]) due to regions of different apparent reflectivity.

In this paper, we investigate two different approaches for discriminating returns from these various surfaces. An algorithm able to identify waveforms as from ocean, leads or floes has been developed. The first two classes will be used for Arctic sea level studies, and the third for estimating freeboard height and ice thickness (not presented in this paper). The classification step is preceded by 'high-level editing' (see Fig. 3) to check that the SGDR confidence flag is 'OK', that the data are in high-resolution (320 MHz) mode and that they present a clear thermal noise region before the leading edge. The results of the two approaches are then compared to generate a hybrid classification algorithm that is used in the production of our Arctic sea level records.

2.2 The Neural Network approach

2.2.1 Building the Neural Network Classification

The first approach implemented in this study consists of a neural network algorithm aiming at classifying the echo shapes and at discriminating exploitable measurements for sea surface height estimation from sea-ice waveforms. Many classification methodologies have already been used to discriminate radar altimetry echoes, from threshold criterion (Laxon [1994]) to Bayesian classifier (Tourneret et al. [2010]). Zhang [2000] presents the advantages of using neural networks for classification activities. He shows that neural networks can provide the estimates of posterior probability required by statistical pattern classifiers. Neural network classifiers directly model discriminant functions (functions enabling the prediction of group membership of a sample based on the values of the input predictive variables) if output values are defined in an appropriate way. Thereby, neural network represents a good supervised classifier. This is why we have chosen this method to perform our waveform classification.

The principle is similar to the classification used in the CNES PISTACH products (Mercier [2010]). A shape class number is assigned to each single 18Hz RA-2 waveform using a neural network algorithm. Following the PISTACH classification procedure, a large number of radar echoes acquired by RA-2 (in Ku-Band only) over different basins (not just the Arctic) and different surfaces (ocean, sea ice, land ice and hydrology) has been analyzed. The purpose is to classify the different geometrical shapes of the echoes and not the different surfaces, even if some links can be made between them. Therefore, the observed echo types have been divided into 12 classes depicted in Fig. 2c. It is important to define not only classes for all echo shapes

of interest, but also for all other waveforms numerous enough to impact the classifier. Even if they do not provide useful information, their identification as a dedicated class number prevents the algorithm from misclassifying them as shapes of interest. Amongst all the identified classes, only classes 1 and 2 are considered here. They respectively represent Brown waveforms observed over open ocean and peaky echoes which are mainly produced by reflection in leads or polynyas but also melt ponds (thin layers of water above a floe). None of the other classes are further exploited here.

The implementation and the parameterization of a neural network are critical steps which determine the classification performance. Several network design parameters must be defined in order to ensure a good learning phase. These include the network size (number of input, output and hidden neurons), the input feature variables, the transfer function (also called the activation function) and the training database definition. We choose to implement a single hidden layer neural network using a sigmoidal function. Cybenko [1989] shows that any continuous function can be approximated by a neural network, having only one internal hidden layer and sigmoidal non-linearity by adapting the number of neurons. Even if a second hidden layer can allow a faster and more consistent response of the network, Bishop [1995] explains that problems can appear during the learning phase due to local minima which make multi-layer networks difficult to use. At the end, the transfer function used for output neurons consists of an identity function providing the probability of a sample to belong to a specific class. To avoid the "curse of dimensionality" detailed by Bishop [1995], we reduced the number of inputs by not considering all the waveform bins as input of the neural network, but instead a set of 7 parameters which fully describe the waveform. These include peakiness, the slope of the logarithm of the trailing edge, the slope of the leading edge, the presence of a peak in the trailing edge (a simple test flag), the average amplitudes and slopes computed in different sub-windows within the waveform.

Then the learning step is performed using a representative training dataset of the real data conditions in order to avoid both over-fitting and under-fitting. To build this database, ~2500 ENVISAT/RA-2 waveforms are preselected over several areas of interest such as sea ice (Arctic and Antarctic), continental ice (Greenland and Antarctic), hydrology (different basins) and coastal zones. Pure open ocean waveforms (with different sea state) are then manually added in order to better represent the standard Brown shape in the training database. Adding these extra ocean data is crucial to have a more representative dataset of the real measurement repartition but it is essential to not exceed 50% in order to not overfit this echo shape and to keep enough data in each other class. Finally, a class number is manually assigned to each waveform and then the learning algorithm is performed. A second dataset is built by randomly picking ENVISAT echoes on the previous regions in order to evaluate the neural network classification performance. The waveform distribution in each class is presented in table 1 for both datasets. The overall misclassification rate of this algorithm is computed from the test database and is presented in table 2 for the 2 classes of interest here: ocean waveforms identified by class 1 and peaky waveforms identified by class 2. The "good identification" percentage corresponds to waveforms identified by manual classification as ocean or peaky, which are also identified by neural network respectively as class 1 or class 2. In the opposite, the percentage of failure corresponds to measurements identified in a different class from that given by the manual classification. The final configuration of the neural network classification is then applied to the entire ENVISAT/RA-2 period.

2.2.2 Classification Results

The geographic distribution of classes assigned by the neural network is illustrated on the left-hand side of Fig. 4. In the North Atlantic virtually 100% of waveforms are deemed to be Class 1 ('ocean'), whereas floes (represented by classes 4, 5 & 6) dominate the central Arctic region. Class 2 echoes (representing leads) occur throughout most of the ice-covered Arctic, but their proportion (for this period) varies from 5 to 20%. These are essential for our effort to track the level of the sea rather than its covering of ice. However some regions, such as the coastal part of the Laptev Sea and the sector 270°-300°E have few returns from leads. An examination of

Classes	1	2	3	4	5	6
Learning database (%)	28.36	3.86	6.27	3.54	8.36	21.26
Test database (%)	19.85	3.92	3.33	7.6	6.3	25.14

Classes	7	8	9	11	12	16
Learning database (%)	2.94	0.82	10.63	3.33	3.54	7.09
Test database (%)	1.61	0.35	5.79	2.04	14.32	9.75

Table 1: ENVISAT/RA-2 waveform distribution in each classes in the learning and test databases identified by manual classification

Classes	1	2	others
Good identification (%)	95.82	89.3	77.84
Failure (%)	4.18	10.7	22.16
Total (%)	100	100	100

Table 2: Performances of the neural network classifier on the test database for ocean waveforms (class 1), peaky waveforms (class 2) and other classes

the neural network classification for other months of the year (not shown) demonstrates that the ocean classification ('Class 1') follows the known migration of the ice edge, and that in the summer large regions have a proportion of class 2 waveforms exceeding 60%. This is particularly associated with regions of ice melt and break-up. In September, class 2 echoes are located all over the remaining sea covered by sea ice except in the Canadian Archipelago and more particularly around the Sverdrup islands (northwest of Greenland).

The temporal evolution of the classes proportion in the Arctic (above 60° of latitude) is plotted in Fig. 5 over all the ENVISAT period. The percentage of class 1 echoes ('ocean') is plotted in black, the proportion of class 2 waveforms ('leads') in red and the ice floe proportion (classes 4, 5 and 6) is in blue. This time series clearly shows the seasonal variation of each class. Above 60°N, the predominant proportion of the ice is well represented. A peak in the ocean proportion identified by the neural network classifier (in black) in September 2007 is clearly visible and corresponds to the record ice extent minimum (Giles et al. [2012]). Finally, a strong increase followed by as strong decrease of the peaky echoes population (class 2 in blue) is noted at each ice break-up period. This behavior gives us confidence in the lead identification, and especially combined with the non-zero proportion during winter. Of course, it is challenging to separate melt pond waveforms from lead waveforms, and no reliable discrimination between the two currently exists. A significant portion of the strong increase of the peaky echoes fraction in early summer may be due to melt ponds and it is possible that a seasonal bias could impact the lead level during the melting period.

2.3 A Multiple Criteria Approach

2.3.1 Principle of the Multiple Criteria Classification

The second approach does not directly use the waveform data in the classification, but rather parameters that the standard processing has already calculated from the waveforms. The chief diagnostics are the normalised backscatter strength (from the robust ice retracker), σ^0, and the pulse peakiness, P_k (Laxon [1994]). Additional parameters were applied to detect waveforms that cannot be classified reliably. Daily records of sea-ice concentration (SIC) from NSIDC (National Snow and Ice Data Center) are used to provide contextual assistance. This method is designed to only select waveforms that can be confidently classified; thus it is expected and

accepted that a high proportion of waveforms will be left unclassified.

The waveforms are classified into "specular" and "diffuse" by using a double threshold scheme applied to the σ^0 and peakiness P_k values:

	$P_k < 3$	$3 \leq P_k \leq 30$	$30 < P_k$
$SIC = 0$	$diffuse \Rightarrow ocean$		
$0 < SIC < 75\%$			
$75\% < SIC$	$diffuse \Rightarrow floe$		$specular \Rightarrow leads$

Table 3: Description of the double threshold scheme for surface classification

As stated earlier, this approach deems a large proportion of data "unclassified" and these are discarded from further analysis. A further round of editing is then applied according to whether the classified waveforms can now allow a good fit to the appropriate waveform model. The main criteria correspond to the position of the waveform within the window, specified by the position of the track point (TP) (see Fig. 2 and section 3), the maximum power in an individual waveform bin (P_{max}) and measures of how well a normalised version of the waveform corresponds to its appropriate fitted model (e.g. fit_{LE} , fit_{TE}, for the r.m.s. fit along the leading / trailing edge respectively). The further requirements for selection are then:

OCEAN: $43 \leq TP \leq 47$ & $fit_{TE} < 0.18$
FLOES: $44 \leq TP \leq 46$ & $fit_{TE} < 0.3$ & $width_{LE} < 1.0$
LEADS: $44.5 \leq TP \leq 46.5$ & $P_{max} > 50$ & $\sigma^0 > 25dB$ & $fit_{TE} < 0.015$ & $P_{tail} < 0.27$

where TP is the output of the retracking algorithm described in Section 3, and $width_{LE}$ is the difference between retracking bin numbers estimated by the using the offset-centre-of-gravity (OCOG) retracker [Wingham et al. 1986] using the 25% and 50% thresholds.
The threshold values were selected experimentally from numerical analysis of waveforms identified manually as unreliable, diffuse or reflective. For each classification, the mean values and deviations of the parameters above were estimated and applied in the threshold selection. The parameters were fine-tuned to minimize the chance of misclassifying unreliable waveforms and introducing errors into the retracking process.

2.3.2 Validation of Multiple-Criteria classification

We validate the waveform classification achieved by the multiple criteria system by using optical data from the Medium-Resolution Imaging Spectrometer (MERIS). An algorithm for sea-ice detection has been developed for MODIS (Willmes, S. and Heinemann [2015]), but here we choose to use MERIS for two main reasons: 1) in its Fine Resolution mode it provides data at 300m resolution commensurate with the waveform spacing for the RA-2 altimeter; 2) it is on the same satellite platform (ENVISAT), which means that the two sets of observations are simultaneous. These are both critical points, as the individual leads are of the order of hundreds of meters across, and features can move several kilometers per day.

An optical classification into "ocean", "floes" and "leads" is particularly challenging in that a fourth possibility exists: "clouds". These are hard to distinguish simply from ice floes. Simple methods based on thresholds of optical parameters (analogous to the ones used to successfully separate water from non-water pixels) do not usually provide accurate results. Thus, more sophisticated methods based on a combination of neural network algorithms, Bayesian statistics and the synergy of optical and thermal sensors (Gomez-Chova et al. [e.g. 2007]; Bulgin et al. [e.g. 2015]; Hollstein et al. [e.g. 2015]) have been recently developed. In this study, cloud

identification was performed using the $IdePix$ algorithm for water pixel classification distributed with the BEAM software suite developed for ESA by Brockmann Consult. The classification is mainly based on the algorithms described in chapter 5 of Bourg [2011], which defines a cloudiness likelihood index ranging from 1 (certainly clear) to 2 (certainly cloud). A cloud mask is then defined by choosing a given threshold. Choosing too low a value results in only open-water pixels being passed, whereas the default value of 1.4 is usually too permissive for our application, with several cloud-covered regions remaining unmasked. Therefore we adopted a value of 1.3, as this allowed us to include cloud-free ice and lead pixels in our validation. However, the technique is not perfect, as evidenced by the red pixels in the southeast portion of Fig. 6b.

This difficulty in finding a threshold that worked for all cases prevented the development of a completely automated validation approach, since the final estimates of the performance of the altimeter classification would have been strongly affected by the inclusion of unmasked cloud pixels. To circumvent the issue, we based the validation on a series of manually identified cloud-free sections within our MERIS images taking note of the Top-of-Atmosphere radiances that had passed the $Idepix$ cloud-clearing. Fig. 6 shows an example of such a section, with Fig. 6c showing the high spatial variability of a region with leads, and Fig. 6d showing the homogeneity characteristic of ice floes.

The selected cloud-free MERIS images were individually examined to identify the parameters to be used for the optics-based classification of open-water, leads or ice regions. Absolute values of radiance were not a useful discriminator, due to their variations as a function of the solar angle. Thus, we normalized all spectra by the radiance value at 489.88 nm. Spectra from open-water pixels are characterized by a sharp exponential decay from blue to red wavelengths, whereas for leads and ice the decay is more gradual (Fig. 7). (The full-width view from MERIS is composed from 5 different cameras; this normalization also minimizes discontinuities between the different cameras' swaths). However, this normalization does not provide a clear separation of ice and leads, as the envelopes of observations for these two classes overlap. This likely occurs because, despite being characterized by open-waters, the optical signal associated with leads is strongly contaminated by the signal associated with the surrounding ice. In order to distinguish between them we look at the heterogeneity within a 7x7 pixel square, which corresponds to the 2km diameter disk that is the minimum resolution of the RA-2 altimeter over flat conditions. Figs. 6c) and 6d) show that such a segment from a region with leads will have considerable variability between the pixels' relative radiances, whereas variation will be small where ice floes are complete.

Denoting the observed radiance at wavelength x by r_x, we define the normalized radiance at 779 nm as $S_{779} = r_{779}/r_{489}$. [The selection of two widely spaced wavelengths simply gives a measure of the spectral slope, without being sensitive to the low values at 761 nm (oxygen absorption, see Fig. 7) or to the wavelengths above 800 nm that may respond to atmospheric cloud.] Then for each altimeter location, we consider the 7x7 array of MERIS pixels surrounding it, but only retain the match-up if all those MERIS pixels are both land- and cloud-free. The two metrics to be considered for the classification are then the minimum value of S_{779} among those pixels and the range i.e. difference between maximum and minimum values.

A total of 42 MERIS cloud-free sections were identified resulting in a total of 5173 waveforms classified by the multiple criteria method (3776 open-water, 887 leads and 510 ice, respectively). A scatter plot of the two metrics shows the open water to be readily identified by low values for both minimum and range, and the leads having higher values for the range than was the case for the floes. Thus a good correspondence is demonstrated between the optical properties in the 7x7 pixels centered on the nominal location of the altimeter return and the surface-type classification produced by the multiple criteria approach.

2.4 Comparison of approaches and the combined classification

The preceding sections have detailed two very different approaches to classification of waveforms in the Arctic. The neural network approach classifies waveforms among a set of characteristic shapes using geometrical and geophysical parameters associated with ocean and typical specular reflectors. The multiple criteria approach uses numerical diagnostics pertinent to each altimeter return, although clearly some (such as peakiness) do provide information on waveform shape rather than amplitude. In this section, we compare the results of these two approaches and define a hybrid classification, which is used to disentangle ocean and lead returns (representing sea level) from those emanating from the sea-ice. The specified algorithm is then used in the creation of the new CCI sea level record for the Arctic.

Fig. 4 and table 4 show the classifications of the two methods for cycle 13 (January-February 2003), which corresponds to a period when the ice coverage is still growing. In this representation, neural network class 1 (see Fig. 2c) is taken to be indicative of ocean returns, classes 4, 5 & 6 to be typical of ice floes and classes 2 to represent different combinations of specular returns indicative of leads within the ice floe. The corresponding groups for the multi-criteria approach are shown in the right-hand column.

Multiple Criteria	**Neural Network**			
	Class 1	Classes 4-6	Classes 2	Other Classes
Ocean	411740	22354	0	137
Floes	10194	219472	0	1606
Leads	0	74656	145130	1580
Unclassified	52383	2155935	76264	47739

Table 4: Correspondence matrix of the two classification schemes for cycle 13, north of 65°N. Colored cells are related to the hybrid classification: blue for the ocean and cyan for the leads.

There are broad similarities between the two, but one major difference is the amount of data classified. The multi-criteria approach is very conservative, in that for 72% of waveforms it does not provide a definite classification as "ocean", "floe" or "lead", whereas for the neural network only 1.6% of data lie in classes other than 1,2,4,5 or 6. A more detailed examination of Fig. 4a shows a number of "class 1" waveforms in high-latitude regions where open ocean is very unlikely to be present. Fig. 4e, Fig. 4f show that there are also a number of waveforms in class 2 in regions for which the multi-criteria approach has no definitive "leads"; in this case, we feel the characteristic waveform shapes picked out by the neural net are credible for narrow leads that fail to pass the σ^0 threshold in the multi-criteria approach.

With these observations in mind, we define a hybrid classification (see Fig. 3) that captures the benefits of each method. It is intended to select only the most reliable SSH values, as the CCI products are designed for studying long-term changes, rather than mapping individual mesoscale features. Thus, the hybrid algorithm leans heavily on the conservative approach to only use sea level from waveforms that can be unequivocally classified and retracked. (as the ultimate goal is to recover accurate records of sea surface height, it is clearly of no benefit here to maintain many different classes of waveforms if they are not then able to be reliably retracked.) The hybrid scheme thus adopts the multiple-criteria classification for "ocean" and treats as "leads" points either classified as such by the multiple-criteria scheme and class 2 or members of class 2 that were "unclassified" in the multi-criteria approach. This increases the number of credible returns within the interior of the polar cap. However, there is a further editing step after the retracking (see section 4).

3 A NEW ADAPTIVE RETRACKING SOLUTION

3.1 Current retracking strategies

The next and crucial step for extending the sea level estimation in the Arctic Ocean is to process the selected waveforms (corresponding to "lead" returns) in order to estimate the altimeter range : this is the role of the retracking step. Altimeters are designed to measure reflections from open ocean which produces waveforms that conform to the Brown model for which retracking techniques are sufficiently mature to provide high accuracy range estimates over open ocean (Amarouche et al. [2004]).

However, in the case of "lead" waveforms which have a peaky shape, the information is carried by only 3 or 4 bins (unlike for ocean returns that spread over all range gates of the waveform), and to retrack accurately those waveforms becomes more difficult. The most widespread retracking technique used to estimate the altimeter range from quasi specular waveforms is the empirical approach. In the case of sea ice regions, and more specifically for the retracking of lead waveforms, Laxon [1994] and Peacock and Laxon [2004] used a threshold retracker which extracts the altimetric range corresponding to 50% of the maximum value of the waveform. Another threshold retracker proposed by Bamber [1994] considers a fraction of the center of gravity of the echo instead of the maximum value. All threshold retrackers are extremely robust and provide an estimate of the altimeter range whatever the waveform shape. However, several limitations are intrinsically linked to this kind of algorithm.

Firstly, these algorithms are purely empirical with no physical bases regarding the radar reflection on the water surface, the antenna gain pattern, the point target response of the instrument (PTR). Over ocean surfaces, it is well known that the altimeter impulse response strongly impacts the values and dependencies of the different estimated parameters, as shown by Thibaut et al. [2010]. This impact is even stronger in the case of the steep leading edges found on specular echoes: the point target response commonly approximated by a $sinc$ shape spreads and shifts the leading edge of the echo, and the position of the retracked point corresponding to a given threshold is dependent to the width and position of the PTR. Secondly, the same threshold cannot be considered for all altimeter missions and has to be adapted with respect to the mission and instrument characteristics (orbit height, antenna aperture, frequency band, etc. ...). Not considering that point makes it very difficult the ability to compare estimates between several missions. Moreover, not accounting for the instrumental Point Target Response is obviously problematic when trying to observe climate signals over the ice cap without considering that the instrument characteristics can itself vary with time (ageing of the electronic components). Finally, the discrete rendition of waveforms means that a small lateral shift of the sampling positions (determined by the relative distance between the satellite and the sea surface knowing moreover that the range gate width is about 47 cm) changes the apparent shape of the waveform (depending where the sampling points are on the peaky echo) (see Fig. 9a). Using a simple threshold method may be the cause of intrinsic errors of the order of 10 cm.

Another retracking approach for quasi-specular echoes has been proposed by Giles et al. [2007]. This retracking solution is based on fitting an empirically derived model specific for peaky waveforms. The empirical function is described by a combination of three functions: a Gaussian for the leading edge, an exponential decaying function for the trailing edge and a polynomial function which links the previous two functions. With this definition, the retracked point corresponds to the maximum of the Gaussian, which means that the radar cross-section is assumed to be a single specular point (Hausleitner et al. [2012]). This definition is not in accordance with Drinkwater [1991] who explains that the calm water within the lead cannot be totally smooth (without any roughness). Fig. 9b shows an indicator of the asymmetry found within ENVISAT waveforms over leads. It is a histogram of the range difference obtained by an OCOG retracker and a simple 50% threshold. The distribution is not fully symmetric because there is, in general, a weak tail to the specular echoes. Again, in this solution, instrumental characteristics (antenna gain pattern and point target response) are not taken into account with

the consequences already explained.
All these points considered, the point corresponding to the lead water surface cannot be where the backscattered power is maximum, but must be located in the leading edge in accordance with the Brown model description.

To conclude, both empirical and Gaussian retrackers are not adapted to process open ocean waveforms. We must recall at this point that they have been designed originally by people working on freeboard height and sea ice thickness, not by people concerned by sea level. A very last point to be discussed is that using these retrackers will not guaranty any estimation continuity with deep ocean measurements which are determined using dedicated retrackers such as the MLE-3 or MLE-4 (Amarouche et al. [2004]) for all conventional missions. Such retrackers are of course based on physical models inherited from Brown's original work. Thus biases may exist between lead and ocean retrackers, which is not desirable when computing complete maps of Arctic sea level. Biases may thus exist (and depend on significant waveheight) due to specific and inhomogeneous "lead" and "ocean" retrackers, which is fully not satisfying when computing complete maps of Arctic sea level.

3.2 A modified Brown model

We describe here a solution accounting for a physical model directly derived from the Brown model (Brown [1977]) but flexible enough to fit peaky lead waveforms (Fig. 2b) as well as diffuse ocean waveforms (Fig. 2a) ; this has been first proposed in Jackson et al. [1992], Brown [1977] (but neglected for simplification), Amarouche et al. [2010] and Amarouche and Vernier [2011].

The double convolution defined by Moore and Williams [1957] and Barrick [1972] describing the radar return as a function of time is written as:

$$S(t) = FSSR(t) * PDF(t) * PTR(t) \quad (1)$$

where:

- $FSSR$ is the flat sea surface response
- PDF is the surface elevation probability density function of scattering elements
- PTR is the radar altimeter point target response

As described by Drinkwater [1991], the effective reflective surface of a radar return from sea-ice plus lead is dramatically reduced to the area of the lead, as the reflective surface of calm water will dominate the surrounding sea-ice return. This effect corresponds to a variation in σ^0 with incidence angle as detailed in Kurtz et al. [2014]. In the Brown formulation, this dependence has been ignored because the model is dedicated to the open ocean with non null sea surface roughness. The author explained that this is a reasonable assumption knowing that the range of incidence angle considered in the case of satellite altimetry is small, and the σ^0 variation with the incidence angle is very low near the nadir. But what is valid for open ocean is no more true for sea ice leads with quite null roughness conditions.

Numerous σ^0 modeling exist in the literature (Barrick [1984]; Jackson et al. [1992]) mainly based on cosine powers and negative exponential. All these modeling assume a Gaussian and isotropic distribution of surface slopes. Amarouche et al. [2010] chose to use a formulation similar to geometrical optics:

$$\sigma_0(\theta) = \sigma_0(0) \exp\left(-sin^2(\theta)/mss\right) \quad (2)$$

with θ being the incidence angle (with respect to the nadir direction) and mss the mean-square surface slope of the dominating reflective surface in the altimeter footprint.

The antenna gain pattern is formulated using an exponential in the Brown model (Brown [1977]). Using (2), the FSSR given in Hayne [1980] can be written without mispointing as:

$$FSSR(t) = A\exp(-\delta t)I_0(0)U(t) \tag{3}$$

with

$$\delta = 4c/(\Gamma h) \tag{4}$$

in which $\Gamma = 4\gamma mss/(4mss + \gamma)$, c is the speed of light, h is the satellite altitude and γ is related to the antenna beamwidth parameter defined by Brown [1977]. In Eq. (3), $I_0(0)$ is a modified Bessel function and $U(t)$ is the unit step function. The double convolution between this new FSSR, the altimeter impulse response (assumed to be Gaussian in the Brown model) and the surface elevation probability density of scattering surface elements is then computed in a similar way as in Hayne [1980]. The resultant model can be formulated accordingly to the Hayne model by replacing γ by Γ. The final formulation (without mispointing) is the following:

$$S(t) = \frac{A\sigma_0}{2}\left[1 + erf\left(\frac{t - \tau - \frac{4c}{\Gamma h}\sigma_c^2}{\sqrt{2}\sigma_c}\right)\right]\exp\left[-\frac{4c}{\Gamma h}\left(t - \tau - \frac{2c}{\Gamma h}\sigma_c^2\right)\right] + N_t \tag{5}$$

where σ^0 is the ocean surface backscattering cross-section at normal incidence, τ is the epoch and N_t is the altimeter additive thermal noise. The σ_c is the composite sigma defined by $\sigma_c = \sqrt{\sigma_s^2 + \sigma_p^2}$ where σ_s is the rms height of the specular points relative to the mean sea level and σ_p is related to the point target response width at -3 dB.

In this formulation, the Γ parameter is related to the surface roughness and for very high mss values (approaching infinity), the Hayne model is retrieved (in this case $\Gamma = \gamma$). For lower values, this parameter notably impacts the slope of the model trailing edge, enabling the fit of peaky echoes as shown in Fig. 10.

Another effect highlighted by the dashed curve in Fig. 11 is the reduction of the model power. In fact, both models have been computed using the same σ^0 distributed over the entire waveform footprint. The returned power is thus proportional to the effective reflective surface. This means that for a similar returned power the corresponding backscatter coefficient will be higher for a smooth surface. Over sea-ice leads, the increase of the maximum waveform power (Drinkwater [1991]; Zakharova et al. [2015]) is consistent with this model feature.
Finally, the position of the water surface level (the key parameter for the estimation in sea-ice leads) in the modified Brown model is still located on the leading edge even in the case of small mss and not at the maximum. It is assumed that sea-ice leads have a non-zero surface roughness, in accordance with the histogram in Fig. 9b.

3.3 The Estimation Process

In our approach, the modified Brown model described by eq. 5 is combined with a Maximum Likelihood Estimator (the same estimation process used for ocean MLE-3 or MLE-4 retrackers) to estimate the ocean geophysical parameters. Four parameters are solved to fit the modified Brown model to the measured waveform: the epoch (τ), the composite sigma (σ_c), the backscatter coefficient (σ^0) and the Γ parameter (related to the mss). Because the modified Brown model has the ability to be equivalent to a classical Brown model (with a strong mss value and its associated Γ parameter), it can also fit well normal ocean waveforms as shown in figure 10a.

Despite that, the retracking algorithm encounters some difficulties in converging for the case of very peaky echoes. The main reason is that the fitting algorithm (MLE) is based on a maximum likelihood criterion which accounts for the speckle noise statistics impacting the waveform (Amarouche et al. [2004]). Over ocean, the speckle noise impacting the 18Hz RA-2 waveforms is assumed to follow a Gamma distribution depending on the number *N* of averaged individual pulses (Ulaby et al. [1982]). The resulting criterion is (Dumont [1985], Rodríguez [1988]):

$$C = cst + N \sum_{k=0}^{K-1} \frac{y_k}{S_k} - (N-1) \sum_{k=0}^{K-1} ln(y_k) + N \sum_{k=0}^{K-1} ln(S_k) \quad (6)$$

with y_k the measured waveform, S_k the model, k the waveform bin number and cst a constant. But in case of radar echoes from sea ice leads, the returned power has been reflected by a very small surface of calm water and the same noise statistic is no more found. Furthermore, the number of non-null waveform bins within peaky echoes is very small and it becomes difficult for the estimator to distinguish signal from the noise. In order to avoid such convergence problems, we choose to adapt the fitting criterion to a pure least square criterion only if the hybrid classification performed beforehand identifies a peaky waveform. The least square criterion used in our adaptive algorithm is:

$$C = \sum_{k=0}^{K-1} (y_k - S_k)^2 \quad (7)$$

Finally, the number of waveform samples considered to adjust the model is reduced to only those carrying information (non-zero points). This results in an adaptive analysis window, whose size depends upon the waveform shape. The waveform class assigned by the hybrid classification described in section 2.5 is used as input of the adaptive retracker. Numbering the waveform bins from 0 to 127, the processing of "ocean waveforms" uses samples 4 to 123 (as for the standard ocean processing) but for "lead waveforms" the last sample used is the eighth point after the waveform maximum. Considering more points is not necessary since their value is zero, but a few null points are required to help the retracker to increase the slope of the model trailing edge through the Γ parameter for specular waveforms. This algorithm strategy is very close to the adaptive strategy used for the RED3 algorithm developed in the frame of the PISTACH project (Mercier [2010]) and the ALES (Passaro et al. [2014]) retracker.These two ocean retrackers focus primarily on the leading edge where most of the information content is, plus a few on the trailing edge decay. However, the adaptive window large enough to include all the bins in the trailing edge of the peaky echo.

Illustrations of the fitting quality of the retracker on class 2 waveforms is shown in Fig. 10 with 3 different examples (b, c and d). Waveform samples are plotted as grey crosses connected by solid grey curves. (The RA-2 waveform contains two additional bins at positions 44.5 and 45.5 on the leading edge; these two bins are used by the fitting process). The black dashed curve represents the modified Brown model resulting from the adaptive retracker with an oversampling factor of 64 (for visualization). These 3 examples show that the fitted model relies on the few non-zero waveform samples and is not limited to these points. The model maximum may be greater than the maximum of the waveform samples. The retracker determines a geophysical solution of the effective reflective surface from the few meaningful waveform samples acquired by the altimeter. As expected for class 2 waveforms, the estimated Γ and σ_c respectively correspond to a very small mss and SWH.

4 Data editing to overcome "hooking"

The preceding sections have covered initial data editing, waveform classification and retracking. In this section, we discuss a final data editing step (see Fig. 3) to remove "uncertain" SSH

retrievals when the altimeter is close to a lead but not directly at its nadir.

In the calculation of sea surface height, there is the assumption that the range recorded onboard the satellite is that to the nearest reflecting surface, which will generally be at nadir. However, the signal from a strongly reflecting lead will dominate the return signal for many consecutive waveforms (Fig. 12). The retracking algorithms tends to follow such a feature leading to large errors in the estimates of surface height, with the distance from such a "bright target" tracing out a hyperbola in the waveform data (Gomez-Enri et al. [2010]). This phenomenon is referred to as “snagging” by Peacock and Laxon [2004] and was given the name “off-nadir hooking” with application to radar altimetry over rivers in Santos da Silva et al. [2010] and Maillard et al. [2015]. Using high-resolution MODIS imagery coincident with an Envisat track, Connor et al. [2009] have shown that reflections from a lead more than 1 km off the sub-satellite track can dominate the signals. The range error related to an off-nadir lead return has been quantified by Armitage and Davidson [2014] using CryoSat-2 SARin data, leading to strong biases in the ice thickness estimation.

To reduce this effect and improve the surface height estimation accuracy, the RA-2 waveforms contaminated by strong off-nadir reflections are automatically edited. The procedure consists in first detecting the waveforms with strong reflection in the nadir direction, and then discarding their neighbouring waveforms affected by off-nadir reflections. Fig. 13a shows the RA-2 Ku-band waveforms over Arctic leads and floes. The variation of waveform intensity along the altimeter track and for the bin position selected at the top of the leading edge is given in Fig. 13b. It is seen that strong nadir reflections produce sharp spikes, which can be automatically discriminated from the rest of waveforms. Individual waveform peaks are compared with a 21-point running average value, with the 5 waveforms immediately before and after such a peak being discarded. Fig. 13d shows that the output of the retracker developed by Giles et al. [2007] is affected by the strong off-nadir reflection effect, with larger range estimates produced in the neighbourhood of spikes. In the output of our adaptive retracker (blue dots) the measurements around spikes are discarded since they are likely to produce biased results. Hence, the developed editing approach improves the retracker accuracy by discarding potentially biased range measurements around strong nadir reflection points.

5 Performance analysis

5.1 Performances in open ocean

The first step for ensuring the sea surface height estimate continuity between the open ocean and the ice-covered ocean consists of validating the adaptive retracker performances in the deep ocean. A full ENVISAT/RA-2 cycle (cycle number 85) has been reprocessed with our new algorithm. Then a new 18Hz altimeter range has been computed from the adaptive retracker estimates with all the necessary corrections detailed in Benveniste et al. [2002] with the aim of being compared with the 18Hz standard altimeter range from the ENVISAT/RA-2 products. Fig. 14 presents a map of the mean differences per boxes of 1°x1° between the new and original altimeter ranges. This map is characterized by a good homogeneity: negligible geographical pattern highlighting a potential wave height dependency is observed. However, a global bias of about -6.28 mm separates the two range estimates with a standard deviation of 2.37 mm. The observed bias can be due to changes introduced in the model but it is not a major concern since the bias is global and the standard deviation of the difference is very small. This result highlights the quality of the adaptive retracker and its good consistency with the standard ENVISAT-RA-2 altimeter range in the open ocean. The few discrepancies observed at the ice edge are due to sea ice measurements that are still present in the open ocean selection. The standard ENVISAT/RA-2 ocean retracker is not dedicated to process sea ice waveforms and consequently provides different estimates than the adaptive retracker on these points.

5.2 Continuity between deep ocean and ice regions

The adaptive retracker has been performed in the Arctic Ocean on waveforms identified as ocean or leads by the hybrid classification. The hooking editing has been applied to echoes from leads as described in part 4. Then the retrieved altimeter range has been subtracted from the ENVISAT/RA-2 orbit to get the sea surface height. The Arctic sea level anomaly (SLA) is built by subtracting the DTU 2013 mean sea surface (Andersen et al. [2015]), wet and dry tropospheric corrections from ERA-Interim, FES 2014 ocean tide correction, earth and pole tide corrections, ionospheric correction from GIM, the inverse barometer correction and finally the ENVISAT/RA-2 sea state bias (SSB) correction but only on ocean measurements. We chose to not apply the SSB correction for leads data because of very small SWH estimates related to specular echoes.

Fig. 15a shows in gray the result of the SLA calculations for the pass 788 of cycle 64, with the background shading indicating where passive microwave data record sea ice concentration (SIC) from OSI-SAF is greater than 50%. Around 77.5°N, there is great variability in derived SLA, but these points are not classified as 'open ocean' by the hybrid classification (which requires SIC=0%, see 3). Data are only accepted as 'leads' if classified by the neural net as class 2 and by the multiple criteria approach as 'unclassified' or 'leads' (the latter of which requires SIC>75%). Given the effective resolution of the SIC gridded fields and our conservative classification approach, we do not find points classified by the hybrid system as 'open ocean' or 'leads' in close proximity. However Fig. 15a shows that there can sometimes be 'unclassified' data showing a near-continuous link of SLA between the other groups. However, our conservative approach avoids the use of such data as their interpretation may be unreliable. In this example the unclassified data appear ∼0.1m higher than the open ocean data, but part of this may relate to the step changes in implementation of SSB and DAC. In the 'open ocean' part, the ENVISAT/RA-2 standard ocean SLA (green crosses) is almost superimposed to the SLA computed with the adaptive retracker which illustrates the agreement shown in the map in Fig. 14. In contrast, if the data believed to be leads are processed with a simple threshold retracker (Ice1 retracker from the ENVISAT/RA-2 products represented by magenta crosses), there is a large offset (∼0.4m in Fig. 15a compared to the SLA computed with the adaptive retracker. Fig. 15b and Fig. 15c show the same comparison on the same track but for 2 different cycles: cycle 66 (March 2008) and cycle 69 (June 2008). The same results are found concerning the SLA computed with the adaptive retracker. The ocean retracker and the adaptive retracker are still consistent, but the offset of the ice1 SLA is varying from one cycle to another. Moreover, the ice1 SLA on leads shows a greater dispersion than the adaptive retracker SLA. These along-track visualizations illustrate the advantages of using a continuous method with the same retracking algorithm to process both open ocean and lead waveforms compared with the computation of a bias between 2 different retrackers. A last point must be noted on Fig. 15c: the lead SLA computed with the adaptive retracker appears to be a few centimeters higher than the ocean SLA. This can be caused by an identification of melt ponds as 'leads' as shown by Armitage et al. [2016]. Discriminating melt ponds from lead waveforms is challenging and there is no melt pond detection in the hybrid classification at this time.

After a simple 3-sigma spatiotemporal editing, gridded maps are built for the entire ENVISAT/RA-2 period with box size of 2° by 1°. One of the computed SLA map is given in Fig. 16 for December 2007. The solid black curve indicates the 50% contour of OSI-SAF sea ice concentration for the corresponding period. The SLA coverage is near complete with a height value in almost every box inside the ice area delimited by the black curve. No significant jump is observed at the transition (from either side of the black curve) which indicates a continuity in the sea level estimate. It may be noted that the strong doming of the Beaufort Gyre is clearly visible and is consistent with Giles et al. [2012].

In order to assess the time consistency of the new Arctic SLA product, we look at the changes between consecutive monthly maps and compare that with another recent Arctic altimetry

dataset produced by DTU (Cheng et al. [2015]). The DTU dataset was built through a re-processing of the corrections and orbits of ERS-1, ERS-2, ENVISAT/RA-2 and CRYOSAT-2 data from September 1992 to October 2014, but with standard retracking applied (Andersen and Piccioni [2016]). In this comparison we consider only the ENVISAT period, and sample both datasets at the same resolution (monthly, 2° longitude by 1° latitude). The internal consistency for each is evaluated by differencing consecutive months and calculating the standard deviation of the changes at each location.
There is much less inter-monthly variability in the newly processed dataset (termed 'CLS/PML', see Fig. 17a) than in the other (Fig. 17b). It indicates that the transition from one map to another is smoother for the CLS/PML product over all the Arctic. Both maps show large variations along the Russian coastlines and in the Foxe Basin which are, however, regions that are seasonally frozen. An interesting feature is the low variation observed in the CLS/PML product at the transition from the open ocean of the Greenland and Norwegian seas to the rest of the inner Arctic Ocean.

Further geophysical analysis has been performed by [Carret et al., 2017] who compare both the CLS/PML and DTU Arctic sea level datasets corrected for steric effects with the observed mass changes recorded by GRACE over the same 2002-2010 time span. For example, this independent regional mean sea level for the Beaufort Gyre shows a correlation of 0.48 with the CLS/PML product, but is -0.33 for the one without retracking. The correlations for the Baffin Bay region are 0.55 and 0.47 respectively.

6 CONCLUSIONS

Radar altimetry is a mature discipline, with sea level records provided on a near-global basis for both climate science and operational applications. ESA's sea level Climate Change Initiative (CCI, see Quartly et al. [2017]) has recognized that the two most challenging domains remaining are the coastal zone and ice-covered seas (principally the Arctic), and has supported research into improvements for both areas. Over the ice-covered Arctic there are two processing challenges: 1) identification of the reflecting surface; 2) robust and accurate derivation of the sea level information.

Although a wide variety of waveforms may be recorded over the Arctic Ocean (see Fig. 2), in practice there are 3 main classes – "ocean", "floes" and "leads". An ocean waveform corresponds to reflections from a homogeneous reflecting surface, with a slow decrease in power after the leading edge (Brown [1977]). Similar waveform shapes are produced when the altimeter footprint is completely covered in rough sea-ice, so daily maps of sea-ice coverage from passive microwave sensors are used to distinguish these two groups. On the other hand, leads or polynyas within the ice cover produce a very different waveform as the reflection from the calm water greatly exceeds that from the surrounding sea-ice to give a specular or "peaky" echo. Laxon [1994] developed the peakiness parameter to characterize such waveforms and used this to discriminate between floes and leads (Peacock and Laxon [2004]). We have built upon this in our two approaches to waveform classification. The classification from the multiple criteria approach has been validated by comparison with optical imagery from MERIS (section 2.3.1), whilst the assessment of the neural network technique will be further detailed in a dedicated paper. For those waveforms classified by both methods, there is a good agreement in their discrimination of waveforms (Table 4), but whilst the neural net classifies almost all records into one of its pre-defined 12 classes, the multiple-criteria approach is more cautious, leaving a higher proportion of data unclassified. As our focus is more on accuracy than quantity of data, a hybrid solution is built combining the two approaches.

The waveforms that have passed the classification (both "oceans" and "leads") are then retracked with the same modified Brown model (section 3.2). This model includes an extra term

Γ that characterises the mean square slope of the reflecting facets of the water surface. For large mss the modified Brown reverts to the Hayne (Hayne [1980]) formulation of the Brown method, and section 5.1 shows that there is negligible bias between the two retracker solutions, providing continuity with the CCI data routinely distributed for the rest of the ocean.

As narrow ice leads can give specular reflections for dozens of contiguous average waveforms, a key aspect of the editing procedure is to only keep the signal corresponding to the brightest return (assuming this to correspond to when the reflecting surface is directly below the satellite) and thus not to use those where the leading edge is caught on some specific feature (known as "hooking" or "snagging"). Whilst this editing removes a significant proportion of the records, those remaining are ones for which there is a high certainty that nadir sea-level is being recorded.

Within the sea level CCI project the retracked ranges from these highly selected ocean and lead waveforms are used to produce monthly maps of sea level on a regular grid. The details of the gridding and the oceanographic applications of these data will be published separately, but they are expected to show new insights into the mesoscale variability complementing Armitage et al. [2017], the seasonal signal in Arctic sea level (due to annual cycles in warming of surface waters, precipitation and riverine run-off, Armitage et al. [2016]) and long-term trends. Giles et al. [2012] had shown a marked spin-up of the Beaufort Gyre up until 2010; our processing of ENVISAT/RA-2 data will need to be combined with those from AltiKa and CRYOSAT altimeters in order to fully map the changes since then.

7 ACKNOWLEDGMENTS

This work has been supported by ESA through the Sea Level CCI (Climate Change Initiative), with additional optical data acquired via the Category-1 proposal, VICTORIA. The authors wish to thank Eero Rinne for the contribution of code and advice, Sara Fleury for discussions on optical recognition of ice, and the National Snow and Ice Data Center for making their data so readily available. Finally the authors wish to thank the French space agency (CNES) who have supported important work on sea-ice classification and dedicated retrackers.

References

Ablain, M., Cazenave, A., Larnicol, G., Balmaseda, M., Cipollini, P., Faugère, Y., Fernandes, M. J., Henry, O., Johannessen, J. A., Knudsen, P., Andersen, O., Legeais, J., Meyssignac, B., Picot, N., Roca, M., Rudenko, S., Scharffenberg, M. G., Stammer, D., Timms, G., and Benveniste, J. (2015). Improved sea level record over the satellite altimetry era (1993-2010) from the Climate Change Initiative project.

Amarouche, L., Jourdain, S., and Deboer, J. R. (2010). Rapport Tâche 2 . 2 : Retracking Spécifique Cellules de pluie et Blooms. *CNES technical report in the frame of SLOOP project*.

Amarouche, L., Thibaut, P., Zanife, O. Z., Dumont, J. P., Vincent, P., and Steunou, N. (2004). Improving the Jason-1 ground retracking to better account for attitude effects. *Marine Geodesy*, 27(1-2):171–197.

Amarouche, L. and Vernier, A. (2011). New methods for retracking altimeter sea ice and ice sheets waveforms. *ESA technical report in the frame of WOOPI project*.

Andersen, O. B., Knudsen, P., and Stenseng, L. (2015). The DTU13 MSS (Mean Sea Surface) and MDT (Mean Dynamic Topography) from 20 years of satellite altimetry. In *International Association of Geodesy Symposia*, volume 12, pages 2006–2007.

Andersen, O. B. and Piccioni, G. (2016). Recent Arctic sea level variations from satellites. *Frontiers in Marine Science*, 3(2296-7745):1–6.

Armitage, T. W. K., Bacon, S., Ridout, A. L., Petty, A. A., Wolbach, S., and Tsamados, M. (2017). Arctic Ocean surface geostrophic circulation2003-2014. *The Cryosphere*, 11(4):1767–1780.

Armitage, T. W. K., Bacon, S., Ridout, A. L., Thomas, S. F., Aksenov, Y., and Wingham, D. J. (2016). Arctic sea surface height variability and change from satellite radar altimetry and GRACE, 2003-2014. *Journal of Geophysical Research: Oceans*, 121 (6):4303–4322.

Armitage, T. W. K. and Davidson, M. W. J. (2014). Using the interferometric capabilities of the ESA CryoSat-2 mission to improve the accuracy of sea ice freeboard retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1):529–536.

Bamber, J. L. (1994). Ice-sheet altimeter processing scheme. *International Journal of Remote Sensing*, 15(4):925–938.

Barrick, D. E. (1972). Remote sensing of sea state by RADAR. In *Ocean 72 - IEEE International Conference on Engineering in the Ocean Environment*, pages 186–192.

Barrick, D. E. (1984). Wind dependence of quasi-specular sea scatter. *Proceedings of the Indian Academy of Sciences - Earth and Planetary Sciences*, 93(2):111–116.

Benveniste, J., Resti, A., Roca, M., Milagro-Perez, M. P., and Levrini, G. (2002). The EnviSat radar altimeter system. *Remote Sensing of the Ocean and Sea Ice 2001*, 4544:71–82.

Bishop, C. M. (1995). Neural networks for pattern recognition. *Journal of the American Statistical Association*, 92:482.

Bourg, L. (2011). MERIS level 2 detailed processing model. *Po-Tn-Mel-Gs-0006*, (8).

Brown, G. (1977). The average impulse response of a rough surface and its applications. *IEEE Transactions on Antennas and Propagation*, 25(1):67–74.

Bulgin, C. E., Eastwood, S., Embury, O., Merchant, C. J., and Donlon, C. (2015). The sea surface temperature climate change initiative: Alternative image classification algorithms for sea-ice affected oceans. *Remote Sensing of Environment*, 162:396–407.

Carret, A., Johannessen, J. A., Andersen, O. B., Ablain, M., Prandi, P., Blazquez, A., and Cazenave, A. (2017). Arctic sea level during the satellite altimetry era. *Surveys in Geophysics*, 38(1):251–275.

Cheng, Y., Andersen, O. B., and Knudsen, P. (2015). An improved 20-year Arctic ocean altimetric sea level data record. *Marine Geodesy*, 38(2):146–162.

Connor, L. N., Laxon, S. W., Ridout, A. L., Krabill, W. B., and McAdoo, D. C. (2009). Comparison of Envisat radar and airborne laser altimeter measurements over Arctic sea ice. *Remote Sensing of Environment*, 113(3):563–570.

Cybenko, G. (1989). Correction: Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems*, 2:303–314.

Drinkwater, M. R. (1991). Ku band airborne radar altimeter observations of marginal sea ice during the 1984 Marginal Ice Zone Experiment. *Journal of Geophysical Research*, 96(C3):4555–4572.

Dumont, J. P. (1985). *Estimation optimale des paramètres altimétriques des signaux radar POSEIDON*. PhD thesis, Institut Polytechnique de Toulouse.

Fu, L.-L. and Cazenave, A. (2000). *Satellite altimetry and Earth sciences: a handbook of techniques and applications*, volume 0. International geophysics series, volume 69 edition.

Giles, K. a., Laxon, S. W., Ridout, A. L., Wingham, D. J., and Bacon, S. (2012). Western Arctic ocean freshwater storage increased by wind-driven spin-up of the Beaufort Gyre. *Nature Geoscience*, 5(3):194–197.

Giles, K. A., Laxon, S. W., Wingham, D. J., Wallis, D. W., Krabill, W. B., Leuschen, C. J., McAdoo, D., Manizade, S. S., and Raney, R. K. (2007). Combined airborne laser and radar altimeter measurements over the Fram Strait in May 2002. *Remote Sensing of Environment*, 111(2):182–194.

Gomez-Chova, L., Camps-Valls, G., Calpe-Maravilla, J., Guanter, L., and Moreno, J. (2007). Cloud-screening algorithm for ENVISAT/MERIS multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):4105–4118.

Gomez-Enri, J., Vignudelli, S., Quartly, G. D., Gommenginger, C. P., Cipollini, P., Challenor, P. G., and Benveniste, J. (2010). Modeling Envisat RA-2 waveforms in the coastal zone: Case study of calm water contamination. *IEEE Geoscience and Remote Sensing Letters*, 7(3):474–478.

Guymer, T. H., Quartly, G. D., and Srokosz, M. A. (1995). The effects of rain on ERS-1 radar altimeter data. *Journal of Atmospheric and Oceanic Technology*, 12(6):1229–1247.

Hausleitner, W., Moser, F., Desjonqueres, J.-D., Boy, F., Picot, N., Weingrill, J., Mertikas, S., and Daskalakis, A. (2012). A new method of precise Jason-2 altimeter calibration using a microwave transponder. *Marine Geodesy*, 35 (suppl.:337–362.

Hayne, G. S. (1980). Radar altimeter mean return waveforms from near-normal-incidence ocean surface scattering. *IEEE Transactions on Antennas and Propagation*, 28(5):687–692.

Hollstein, A., Fischer, J., Carbajal Henken, C., and Preusker, R. (2015). Bayesian cloud detection for MERIS, AATSR, and their combination. *Atmospheric Measurement Techniques*, 8(4):1757–1771.

Jackson, F. C., Walton, W. T., Hines, D. E., Walter, B. A., and Peng, C. Y. (1992). Sea surface mean square slope from Ku-band backscatter data. *Journal of Geophysical Research*, 97(NO. C7):11,411–11,427.

Kurtz, N. T., Galin, N., and Studinger, M. (2014). An improved CryoSat-2 sea ice freeboard retrieval algorithm through the use of waveform fitting. *Cryosphere*, 8(4):1217–1237.

Kwok, R. and Rothrock, D. A. (2009). Decline in Arctic sea ice thickness from submarine and ICESat records: 1958-2008. *Geophysical Research Letters*, 36(15).

Laxon, S. and McAdoo, D. (1994). Arctic Ocean Gravity Field Derived From ERS-1 Satellite Altimetry. *Science (New York, N.Y.)*, 265(5172):621–624.

Laxon, S. W. (1994). Sea-ice altimeter processing scheme at the EODC. *International Journal of Remote Sensing*, 15(4):915–924.

Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S., and Davidson, M. (2013). CryoSat-2 estimates of Arctic sea ice thickness and volume. *Geophysical Research Letters*, 40(4):732–737.

Le Traon, P. Y. (2013). From satellite altimetry to Argo and operational oceanography: Three revolutions in oceanography. *Ocean Science*, 9(5):901–915.

Maillard, P., Bercher, N., and Calmant, S. (2015). New processing approaches on the retrieval of water levels in Envisat and SARAL radar altimetry over rivers: A case study of the S??o Francisco River, Brazil. *Remote Sensing of Environment*, 156:226–241.

Mercier, F. (2010). Coastal and Hydrology Altimetry product (PISTACH) handbook. Technical report, CNES.

Moore, R. K. and Williams, C. S. (1957). Radar terrain return at near-vertical incidence. *Proceedings Of the Institute Of Radio Engineers*, 45(2):228–238.

Morison, J., Kwok, R., Peralta-Ferriz, C., Alkire, M., Rigor, I., Andersen, R., and Steele, M. (2012). Changing Arctic Ocean freshwater pathways. *Nature*, 481(7379):66–70.

Passaro, M., Cipollini, P., Vignudelli, S., Quartly, G. D., and Snaith, H. M. (2014). ALES: A multi-mission adaptive subwaveform retracker for coastal and open ocean altimetry. *Remote Sensing of Environment*, 145:173–189.

Peacock, N. R. and Laxon, S. W. (2004). Sea surface height determination in the Arctic ocean from ERS altimetry. *Journal of Geophysical Research C: Oceans*, 109(7):art. no. C07001.

Prandi, P., Ablain, M., Cazenave, A., and Picot, N. (2012). A new estimation of mean sea level in the Arctic ocean from satellite altimetry. *Marine Geodesy*, 35(July 2014):61–81.

Quartly, G. D., Legeais, J.-F., Ablain, M., Zawadzki, L., Fernandes, M. J., Rudenko, S., Carrère, L., García, P. N., Cipollini, P., Andersen, O. B., Poisson, J.-C., Njiche, S. M., Cazenave, A., and Benveniste, J. (2017). A new phase in the production of quality-controlled sea level data. *Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2017-23*, in review.

Rodríguez, E. (1988). Altimetry for non-Gaussian oceans: height biases and estimation of parameters. *Journal of Geophysical Research*, 93(C11):14107–14120.

Santos da Silva, J., Calmant, S., Seyler, F., Rotunno Filho, O. C., Cochonneau, G., and Mansur, W. J. (2010). Water levels in the Amazon basin derived from the ERS 2 and ENVISAT radar altimetry missions. *Remote Sensing of Environment*, 114(10):2160–2181.

Serreze, M. C., Barrett, A. P., Slater, A. G., Woodgate, R. A., Aagaard, K., Lammers, R. B., Steele, M., Moritz, R., Meredith, M., and Lee, C. M. (2006). The large-scale freshwater cycle of the Arctic. *Journal of Geophysical Research: Oceans*, 111(11):art. no. C11010.

Steele, M., Ermold, W., and Zhang, J. (2008). Arctic Ocean surface warming trends over the past 100 years. *Geophysical Research Letters*, 35(2):1–6.

Thibaut, P., Poisson, J. C., Bronner, E., and Picot, N. (2010). Relative performance of the MLE3 and MLE4 retracking algorithms on Jason-2 altimeter waveforms. *Marine Geodesy*, 33(sup1):317–335.

Tourneret, J., Mailhes, C., Severini, J., and Thibaut, P. (2010). SHAPE CLASSIFICATION OF ALTIMETRIC SIGNALS USING ANOMALY DETECTION AND BAYES DECISION RULE. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, volume 1, pages 1222–1225, Honolulu, HI, USA. IEEE.

Ulaby, F. T., Moore, R. K., and Fung, A. K. (1982). *Microwave Remote Sensing: Active and Passive Volume II: Radar Remote Sensing and Surface Scattering and Emission Theory*, volume 2.

Willmes, S. and Heinemann, G. (2015). Pan-Arctic lead detection from MODIS thermal infrared imagery. *Annals of Glaciology*, 56(69):29–37.

Zakharova, E. A., Fleury, S., Guerreiro, K., Willmes, S., Rémy, F., Kouraev, A. V., and Heinemann, G. (2015). Sea ice leads detection using SARAL/AltiKa altimeter. *Marine Geodesy*, 38(sup1):522–533.

Zhang, G. (2000). Neural networks for classification: a survey. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 30(4):451–462.

8 FIGURES

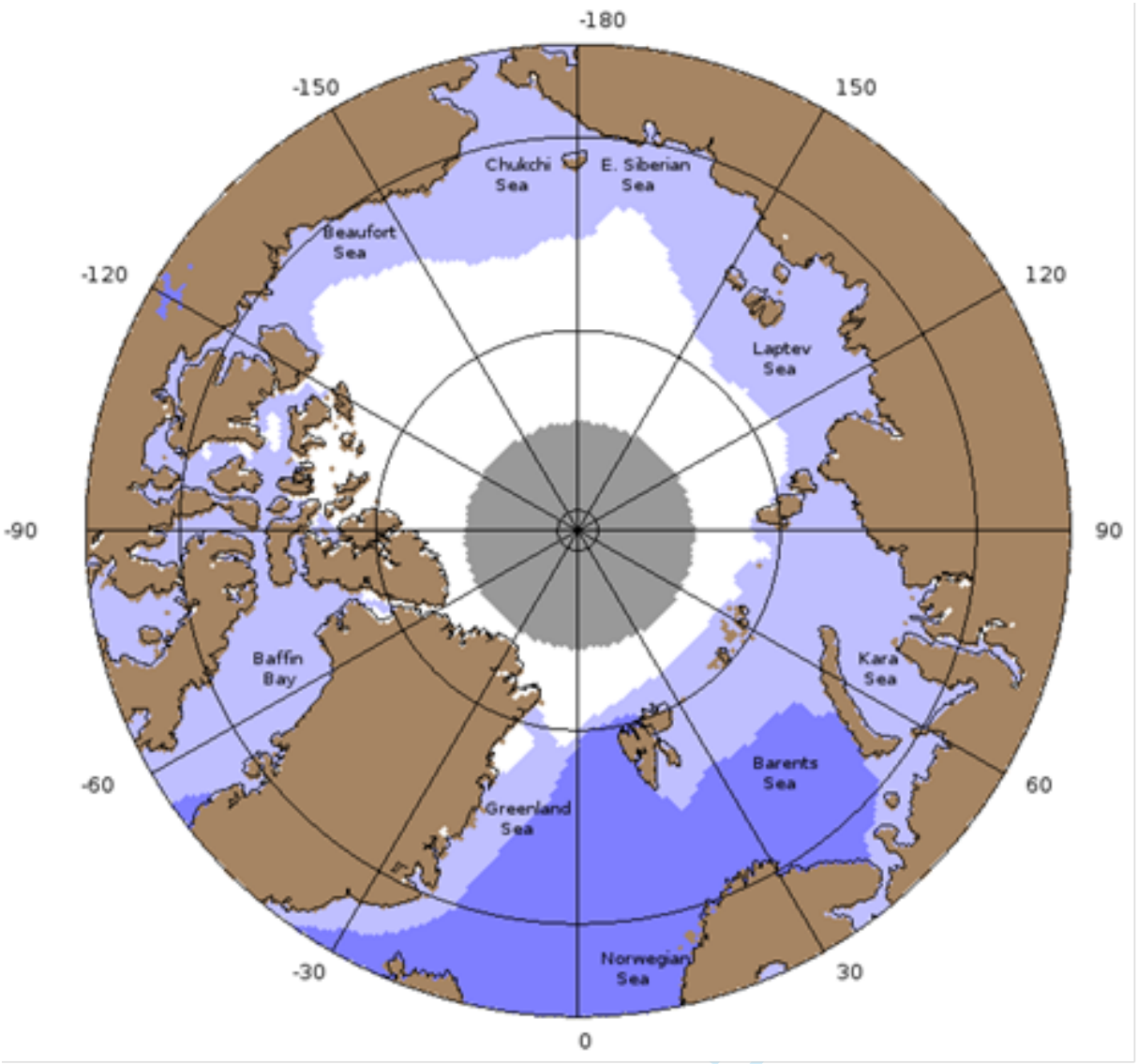

Figure 1: Map of Arctic Ocean, with mean ice extent in September shown in white and the mean ice extension for March in light blue. (Figure derived from 50% sea ice concentrations for March and September from SSM/I data for 1979-2013, provided by NSIDC. The central grey disk marks region for which there is no SSM/I coverage. Individual years will have greater or less coverage than shown.)

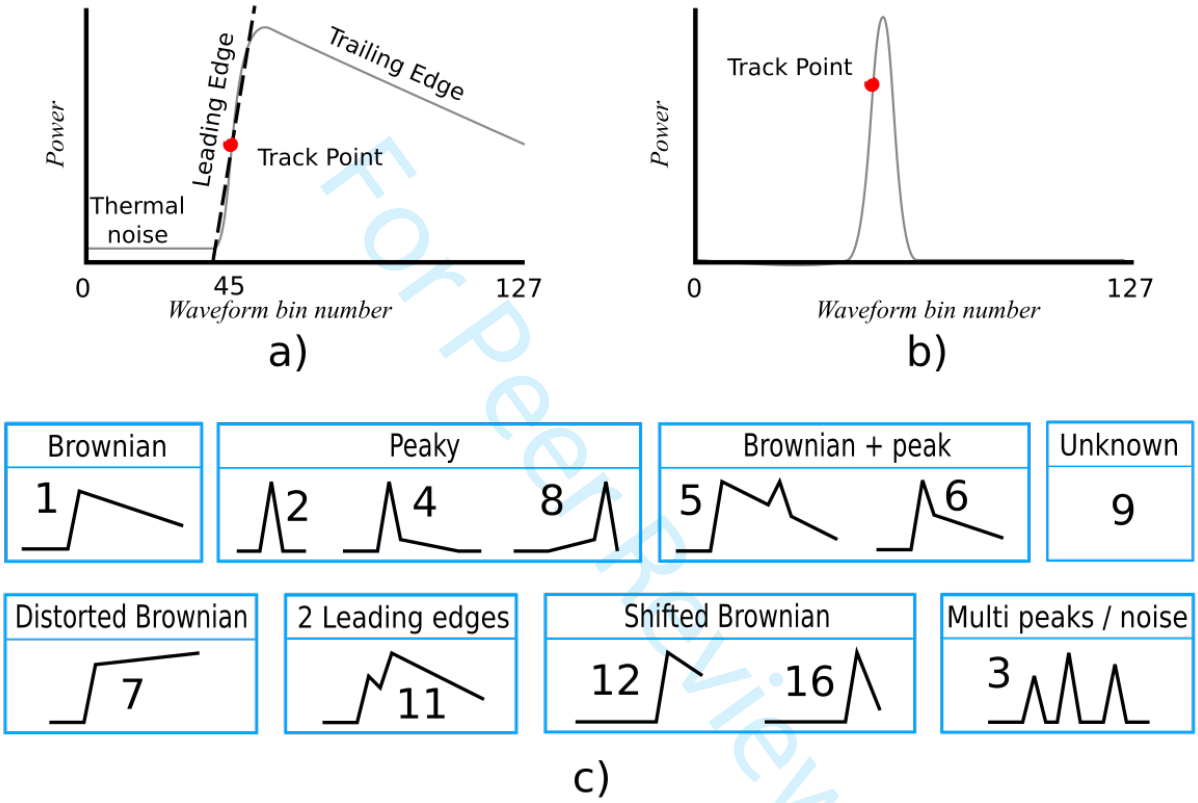

Figure 2: Illustration of the different Arctic waveform shapes. Schematic of a) typical ocean ('Brown') waveform, and b) ice lead waveform, both of which are retracked in this paper. c) The more extensive variety of waveform shapes that may be encountered, and upon which the neural net is trained.

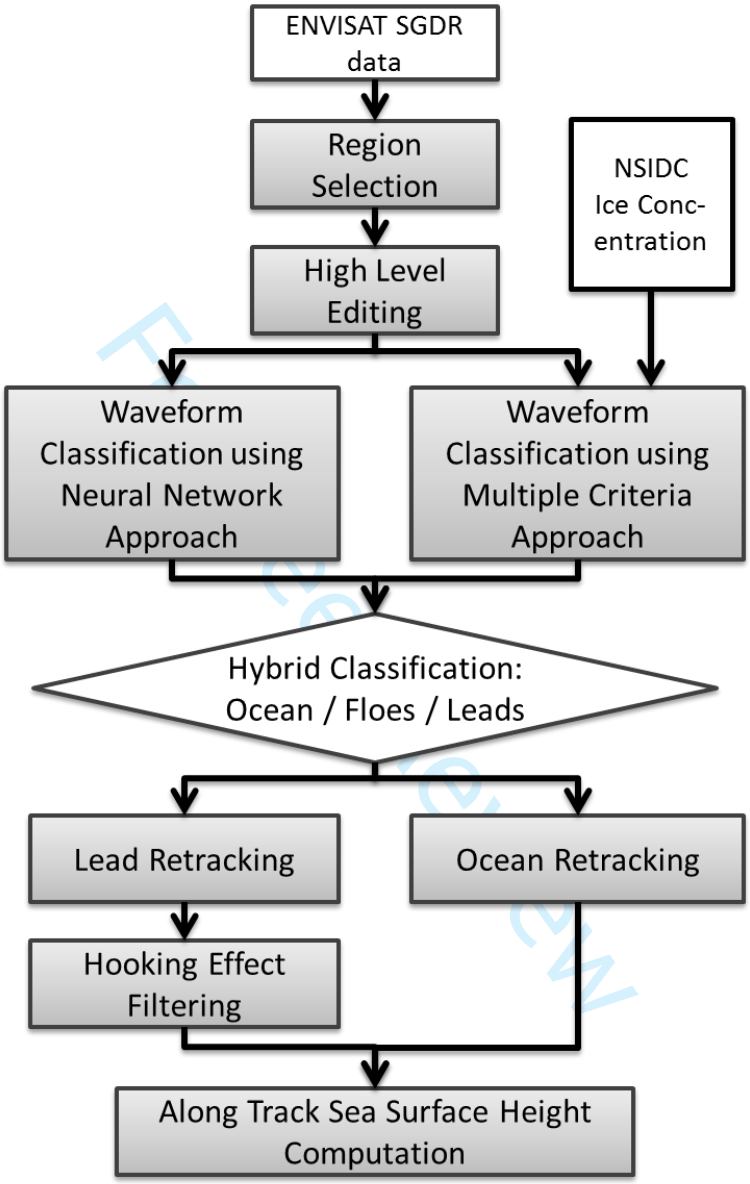

Figure 3: Flowchart describing the data processing for the along track sea surface height computation in the Arctic Ocean

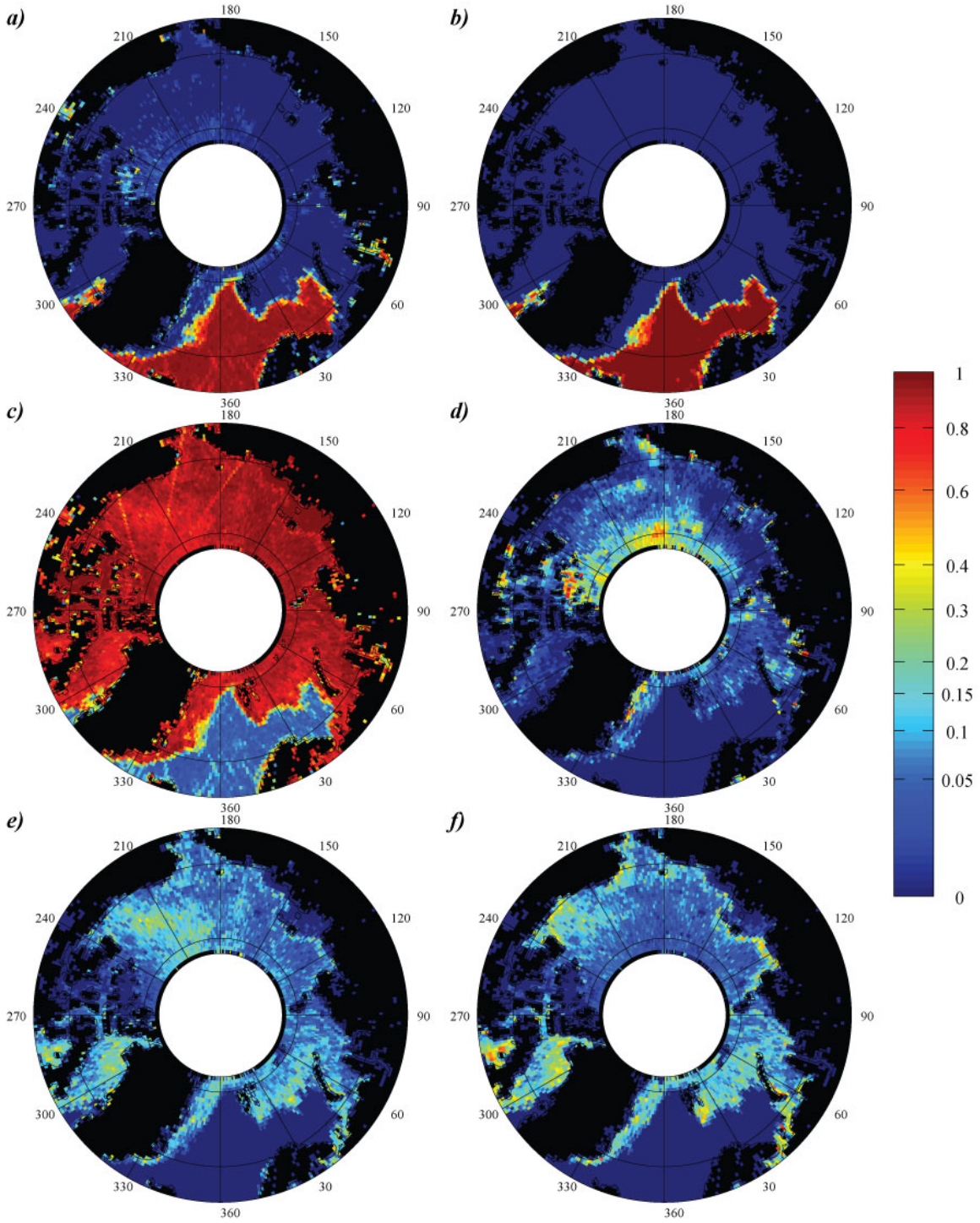

Figure 4: Comparison of percentages of classifications according to neural net (left-hand column) and multiple criteria approach (right hand column). Top row is "ocean" (class 1); middle row is "floes" (classes 4-6); bottom row is "leads" (class 2) colorbar has a non-linear scale to enhance the differences at low percentages.

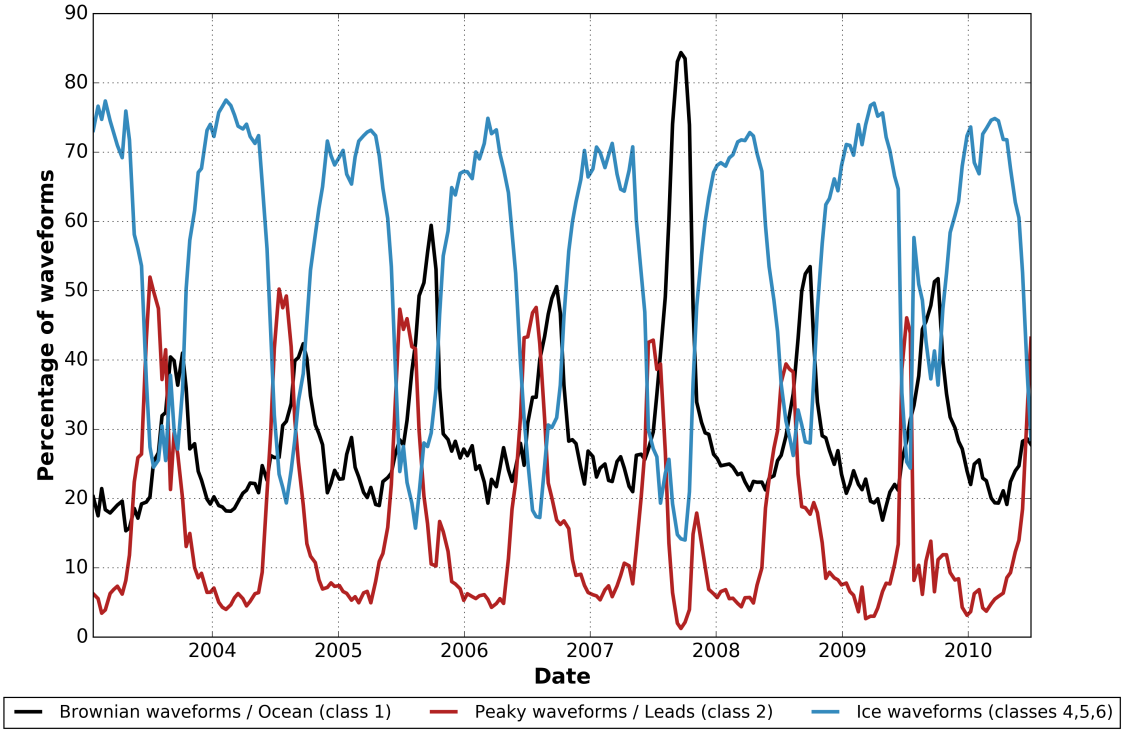

Figure 5: Temporal evolution of the class proportion identified by the neural network algorithm during the ENVISAT period over the ocean above 60 degrees of latitude.

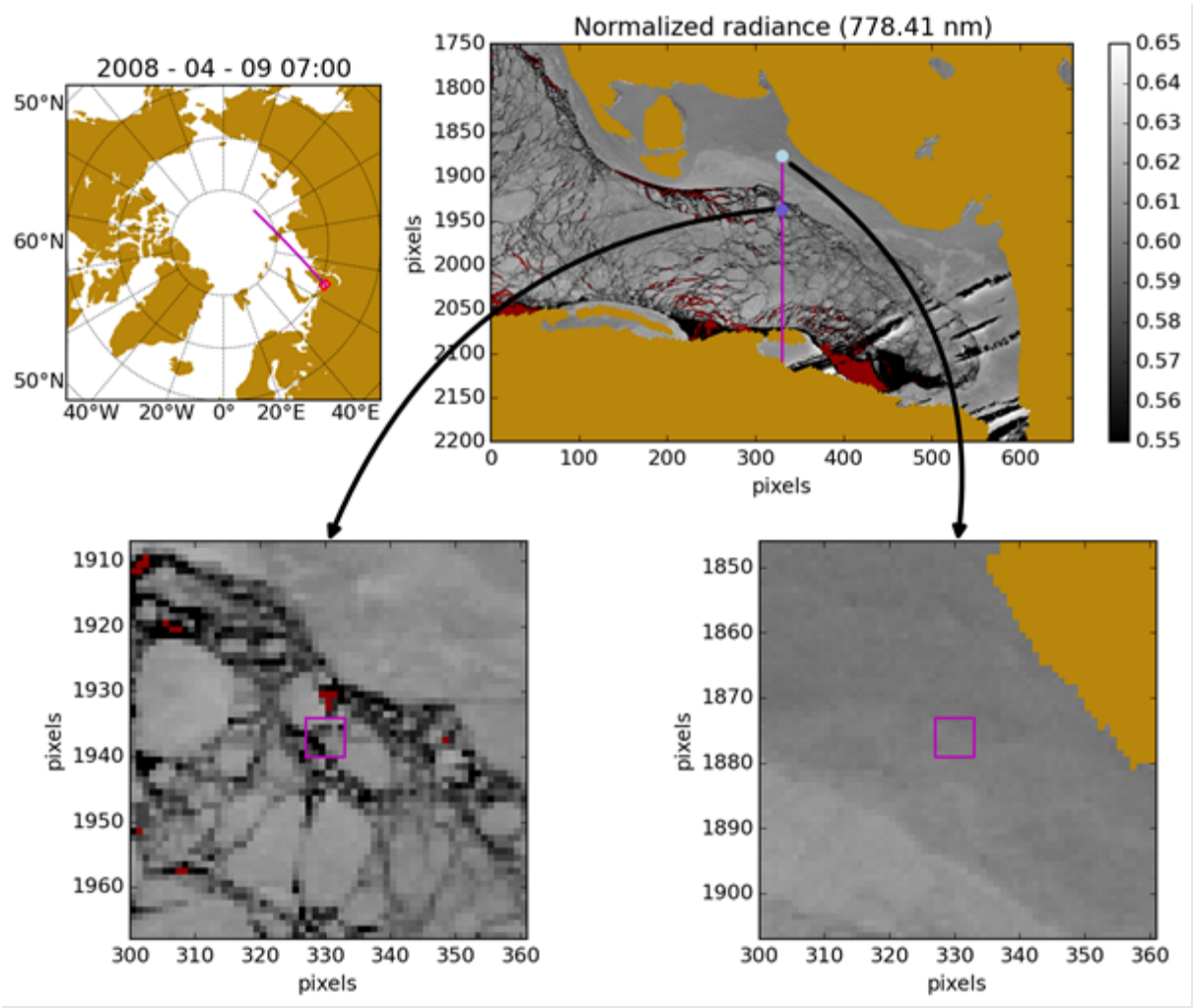

Figure 6: Example MERIS data for 9th April 2008. a) Map showing location of swath. b) Zoom of region at southern end of track, with pink line showing altimeter nadir view, orange indicating land and red marking cloud. (Note unflagged pixels in the bottom right corner of image). c), d) Details of regions of leads and ice floe respectively, showing the greater spatial heterogeneity in the former. (Pink box indicates example regions of 7x7 pixels.

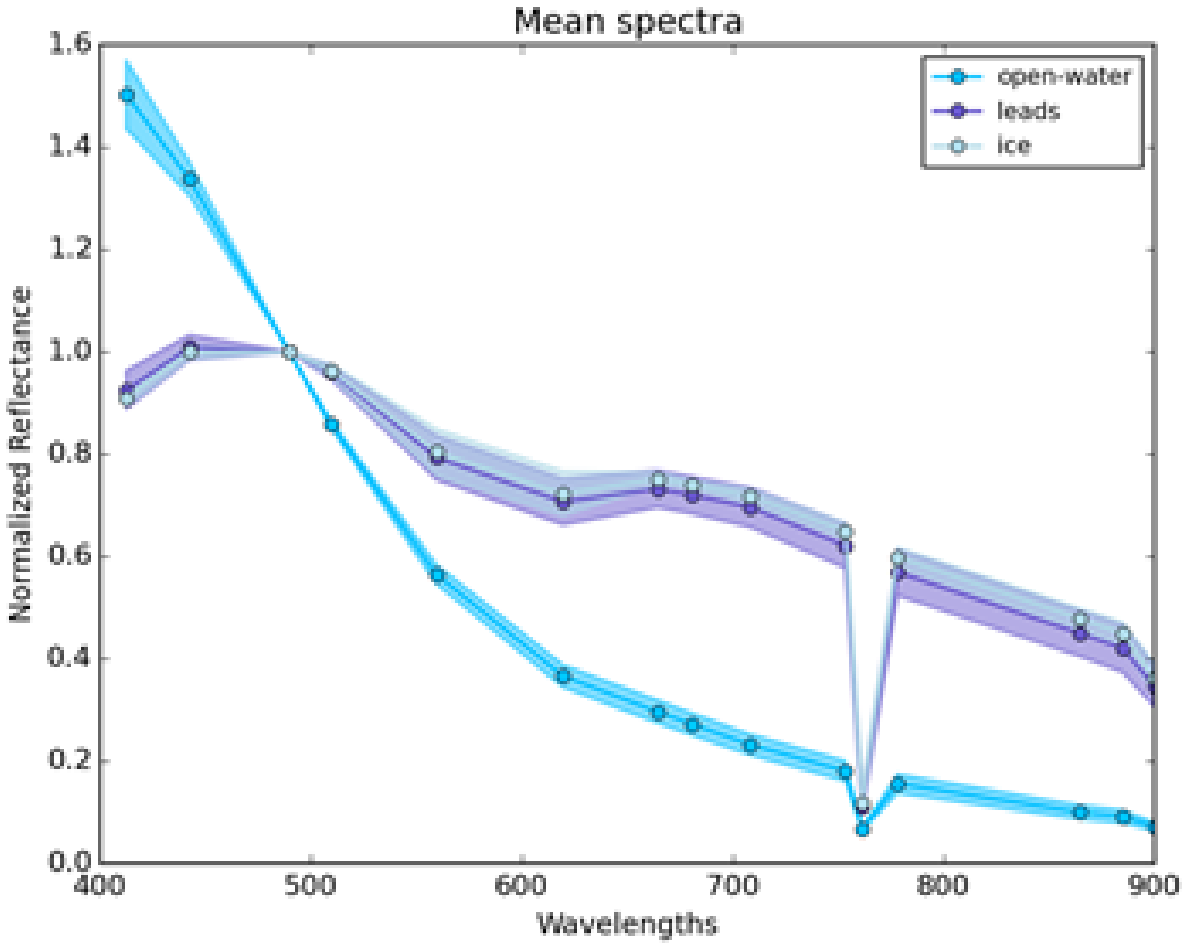

Figure 7: Spectra normalised by value at 489.88 nm, with colouring according to the altimeter classification. Solid line shows the mean, and the shading +/-1 s.d.

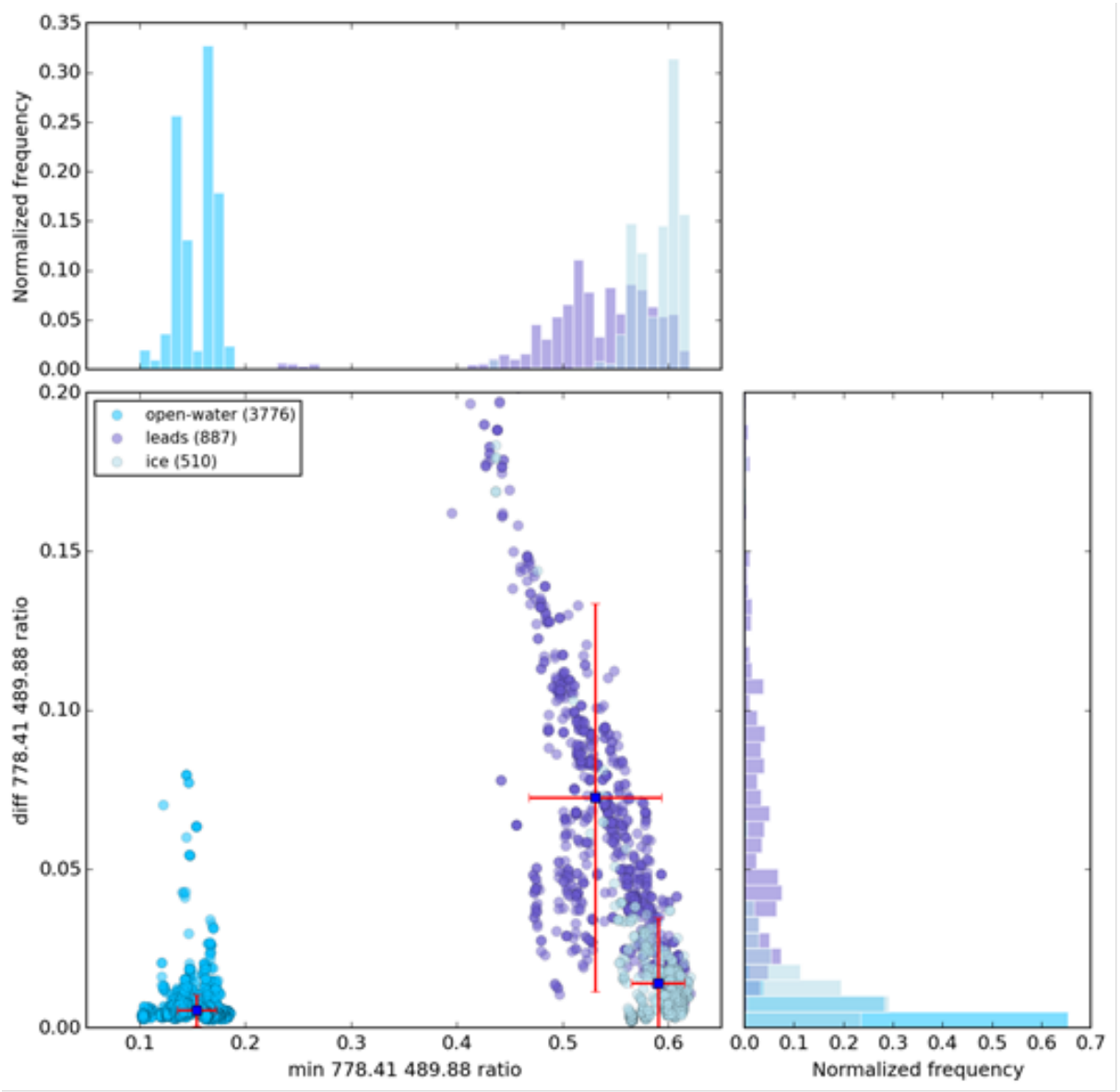

Figure 8: Variation of normalised radiance (r_{778}/r_{490}) within a 7x7 box. b) (bottom left) shows scatter plot of minimum value against the difference of the maximum and minimum values, with colouring according to the altimeter classification. a), c) histograms for the two axes.

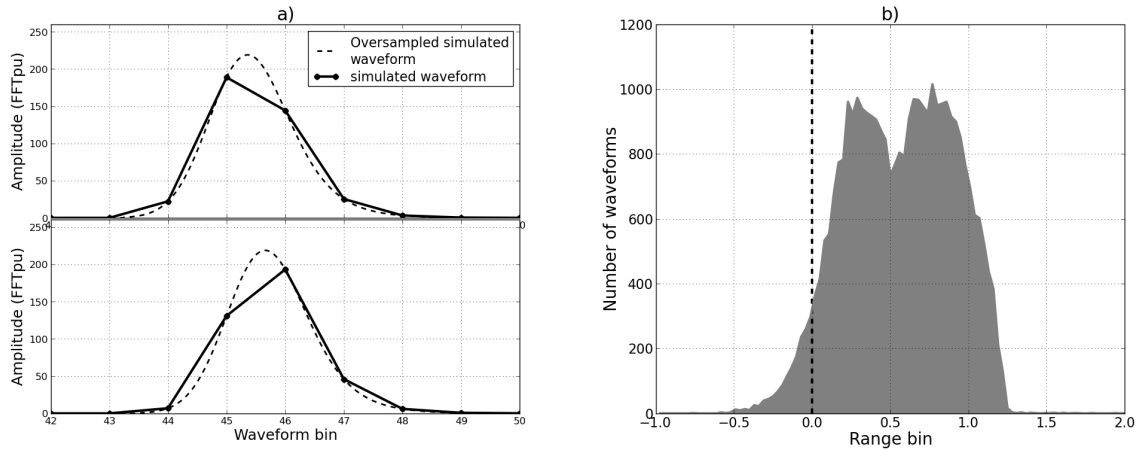

Figure 9: Illustration of the challenge of accurately resolving waveform position for a peaky echo only persisting for a few bins a) Schematic of idealised curve (dashed) and the discrete bins recorded (solid line) for waveform positions separated by one third of a waveform bin. b) Analysis of real ENVISAT peaky waveforms, showing the difference in range bin recorded for OCOG and 50% threshold retrackers.

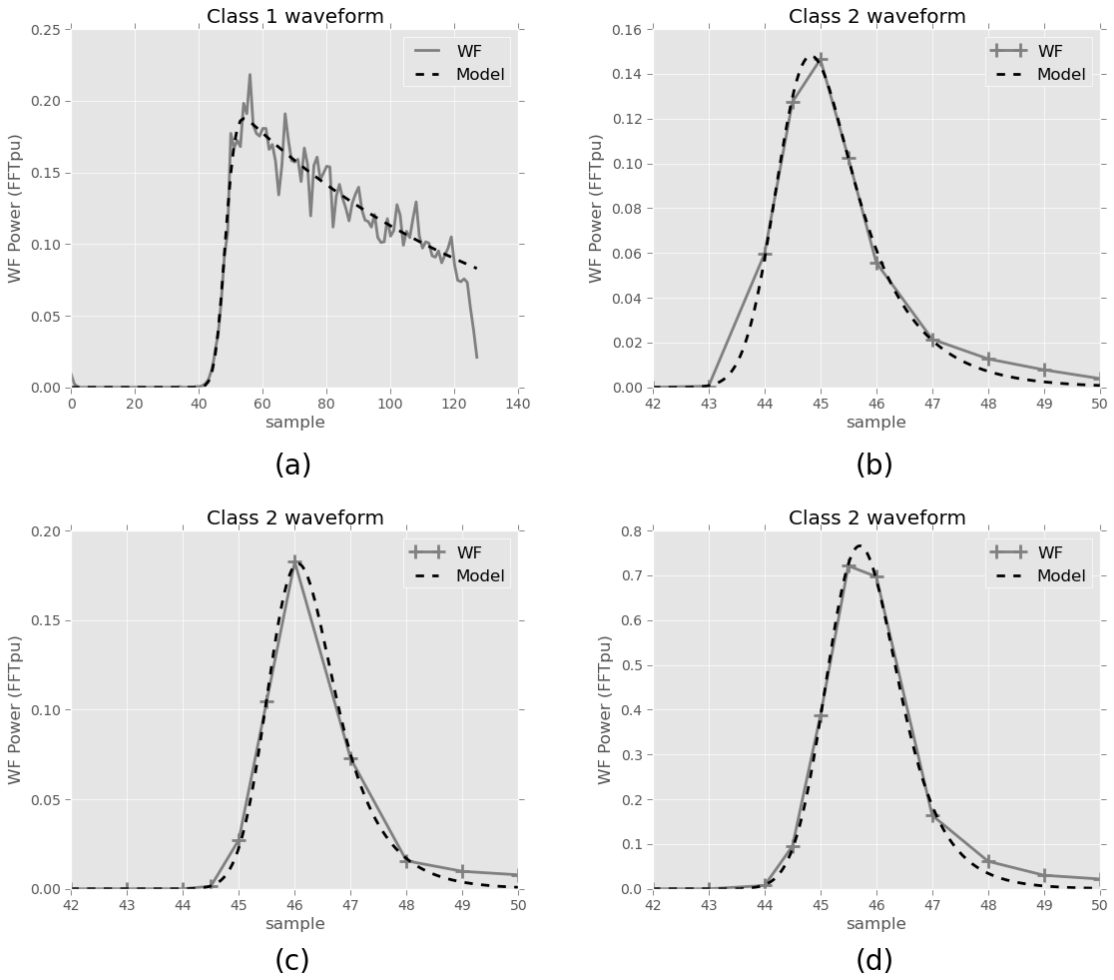

Figure 10: Retracked modified Brown model superimposed on ENVISAT waveforms over ocean (a), over 3 different sea ice lead waveforms (b), (c) and (d) with their associated waveform class assigned by the neural network classification. The 2 DFT samples added by the RA-2 altimeter are considered and are localized at samples 44.5 and 45.5 (starting from 0).

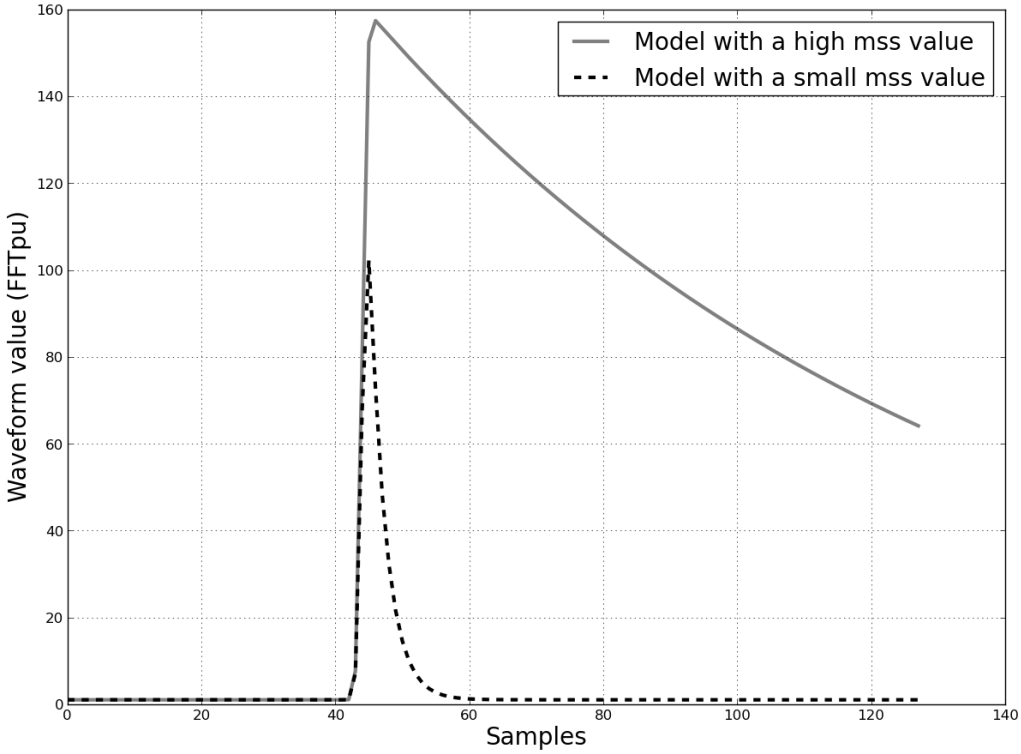

Figure 11: Modified Brown model with a high *mss* value (solid grey); Modified Brown model with a small *mss* value (dashed black)

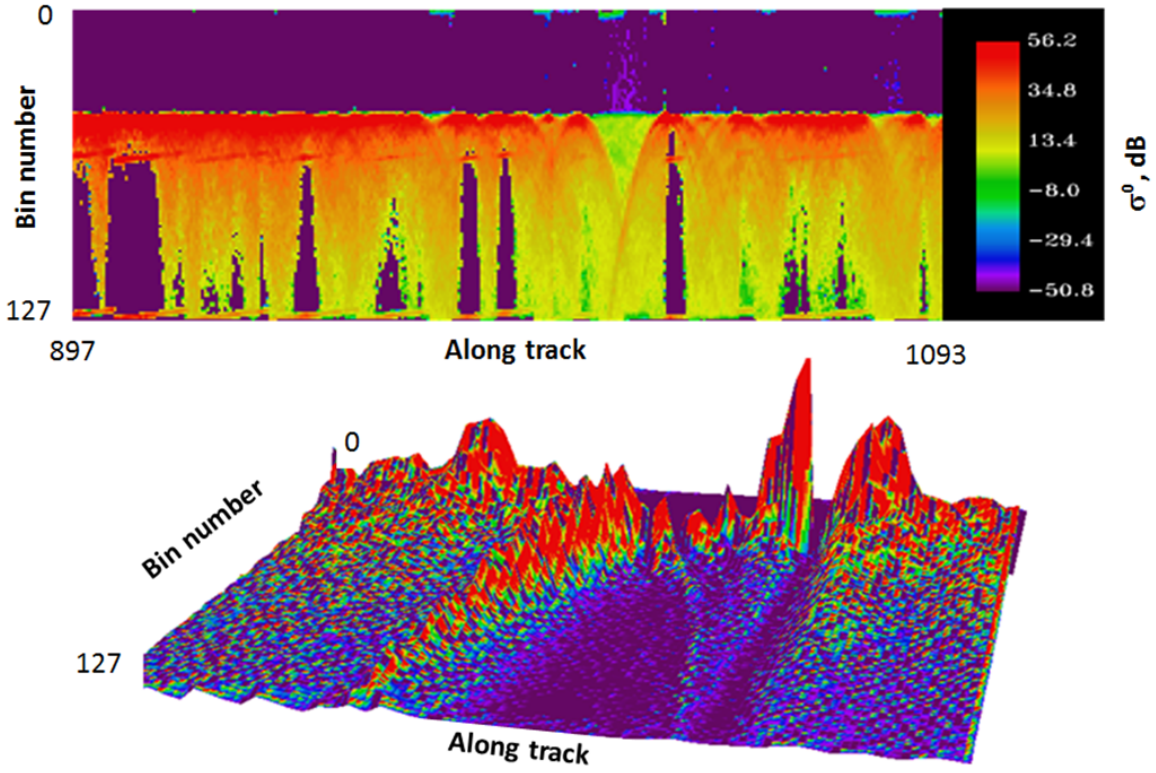

Figure 12: Example of hooking where leading edge follows a hyperbola as altimeter approaches and recedes from a bright specular point. a) Radargram, illustrating power in each waveform bin for a section ∼70 km along track (196 waveforms), with the colour indicating intensity on a logarithmic scale to show the wide range of values. There is minimal power in bins 0-40, with a steep leading edge around bin 45. Strong reflectors dominate even away from nadir, with the delayed responses producing hyperbolae. b) A 3-D representation of the power in each bin for a shorter segment.

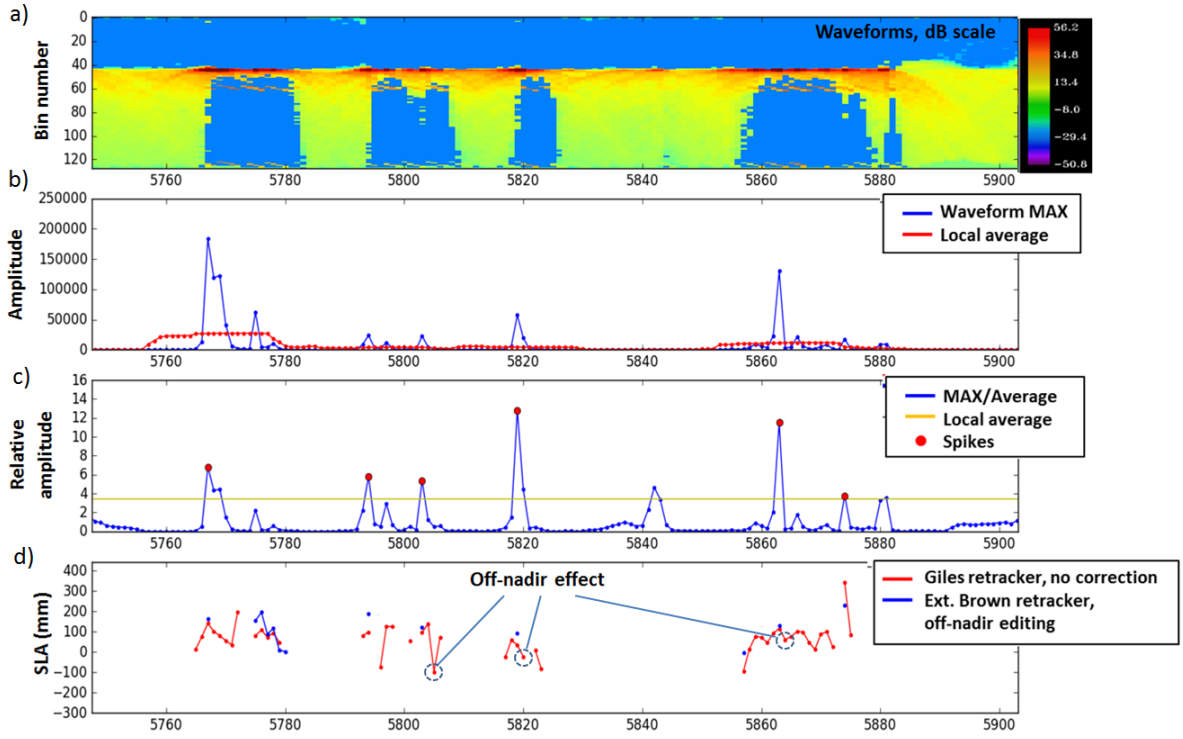

Figure 13: a) Series of 370 ENVISAT RA-2 waveforms, spanning 120 km along track. b) Power value in bin at top of leading edge (absolute units) and result of a running 21-point mean. c) Ratio of actual value to 21-point mean, with a threshold for peak detection at 3.5. d) Sea Level Anomaly calculated using the Giles and the Adaptive retracker. For the latter, the 5 waveforms either side of a "significant peak"are discarded, so the rapid spatial changes associated with hooking are avoided.

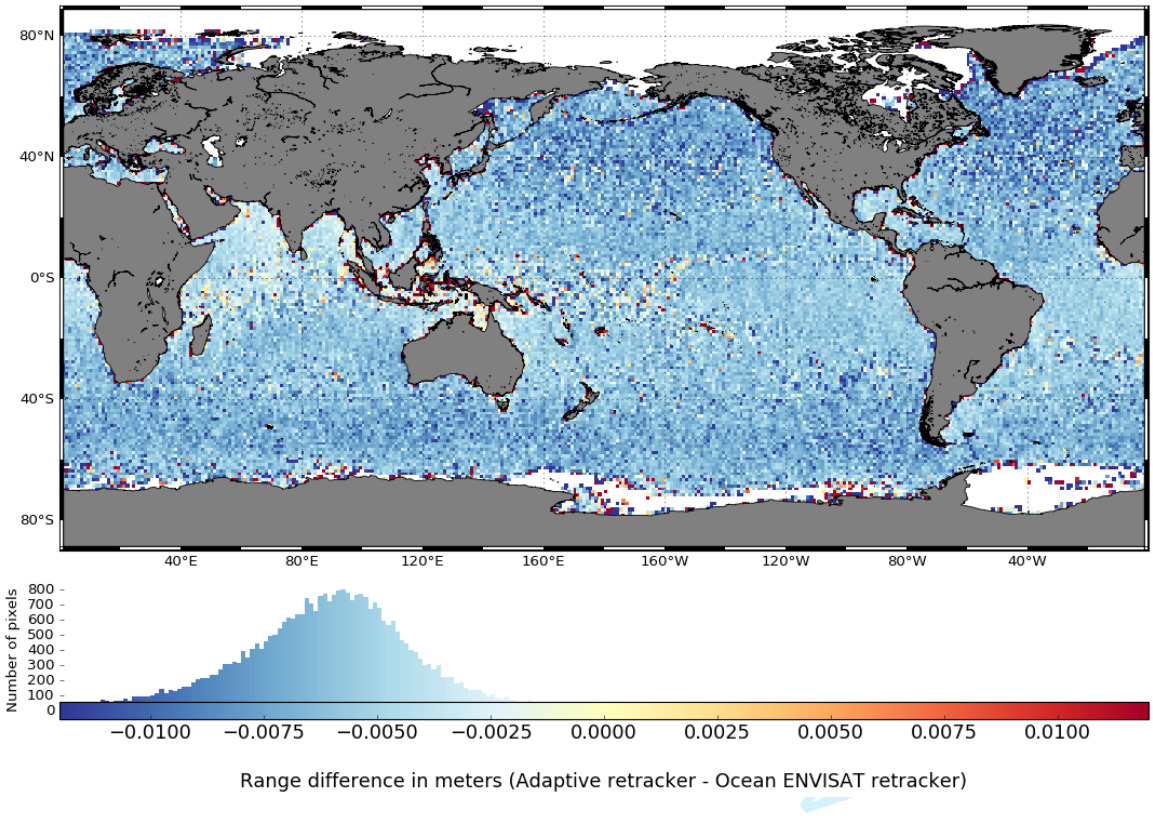

Figure 14: Gridded map of the range difference between the adaptive retracker and the standard ENVISAT/RA-2 ocean retracker over open ocean on cycle 85 (1°x1°). The histogram of the differences is presented at the bottom, which has a mean bias of -6.28 mm and a std. dev. of 2.37 mm.

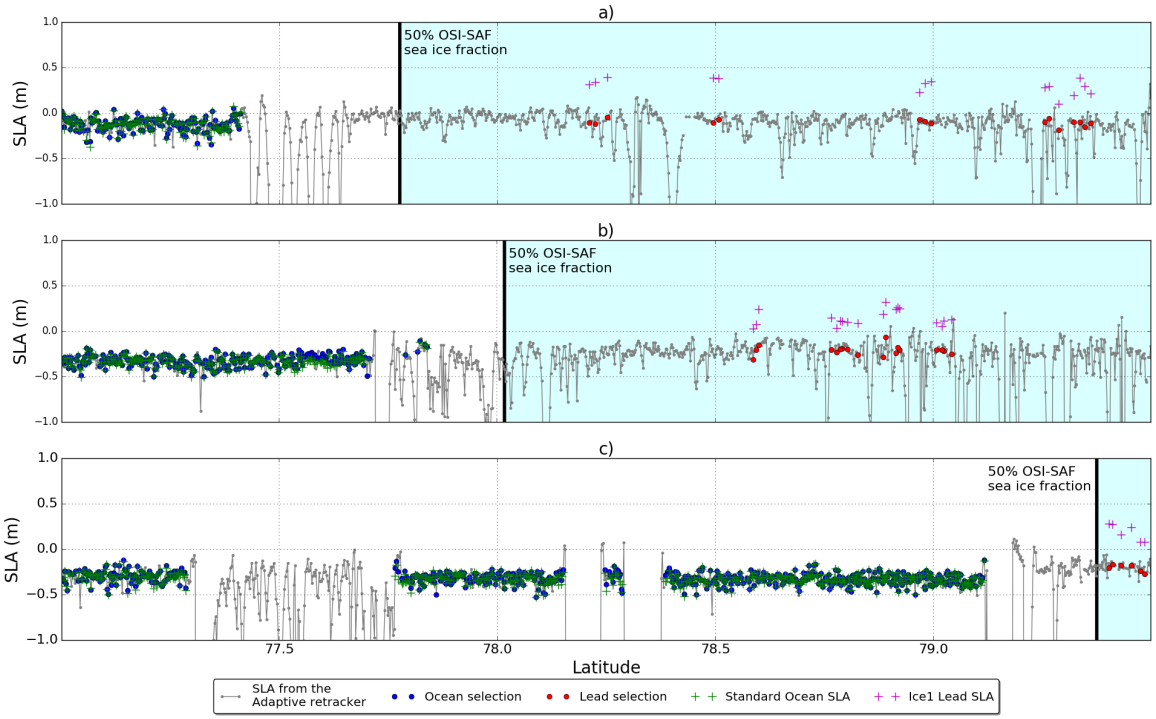

Figure 15: SLA profiles on EnviSat pass 788 for a) cycle 64 (December 2007), b) cycle 66 (March 2008) and c) cycle 69 (June 2008), contrasting the performances of different retrackers. The SLA computed with the adaptive retracker is plotted in gray. Measurements classified as 'ocean' are colored in dark blue, points classified as 'leads' are colored in red. Green crosses represent the SLA computed with the ENVISAT standard ocean retracker on the 'ocean' points whereas the magenta crosses correspond to the SLA computed using the ice1 retracker from the ENVISAT products on 'leads' points. Only points definitely classified as 'ocean' or as 'leads' are used in our analysis. The output of the standard ocean retracker lies close to that of the adaptive retracker, whereas, in this instance, the ice1 retracker for leads gives much higher values.

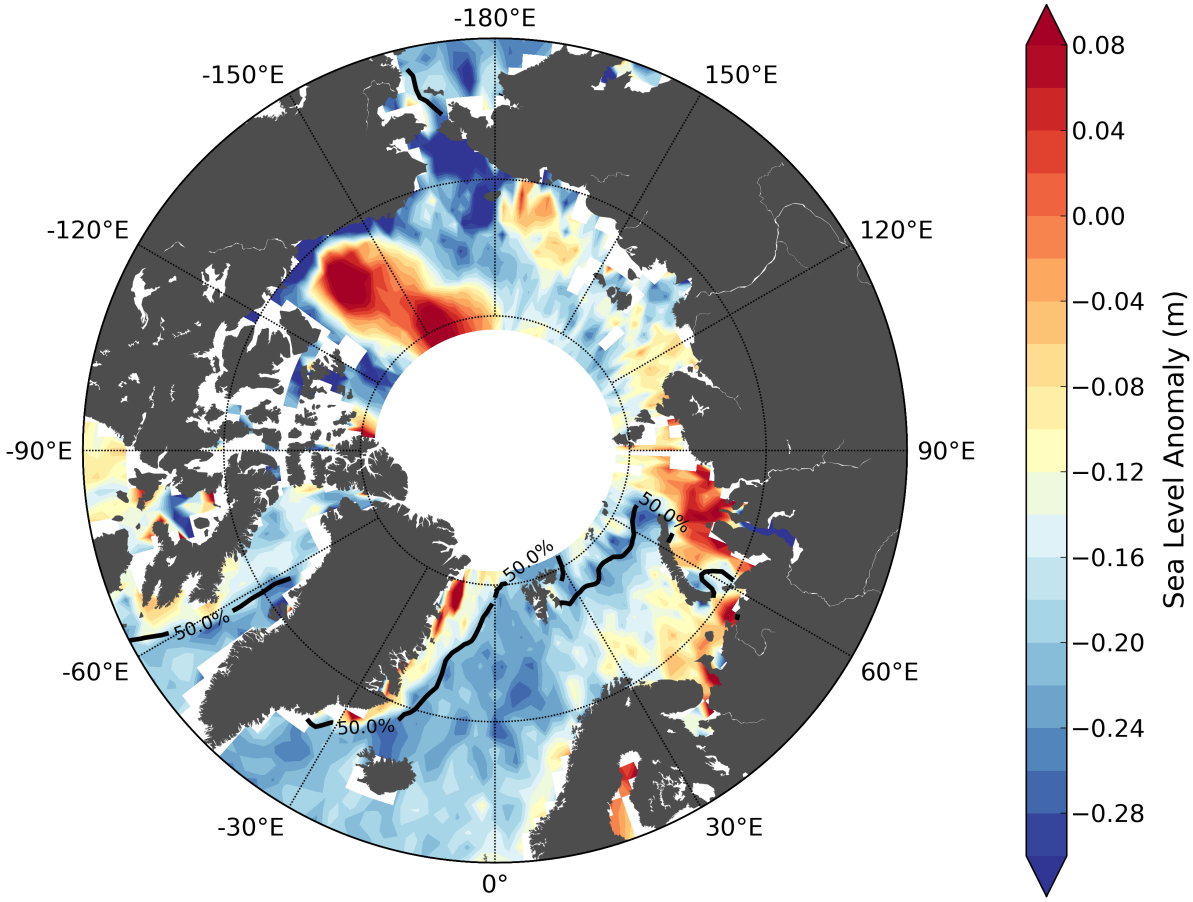

Figure 16: Gridded map of the Arctic sea level anomaly for December 2007 produced by the new retracking solution (box 2°x1°). The solid dark curve is the contour of 50% sea-ice concentration (from OSI-SAF) for the same month.

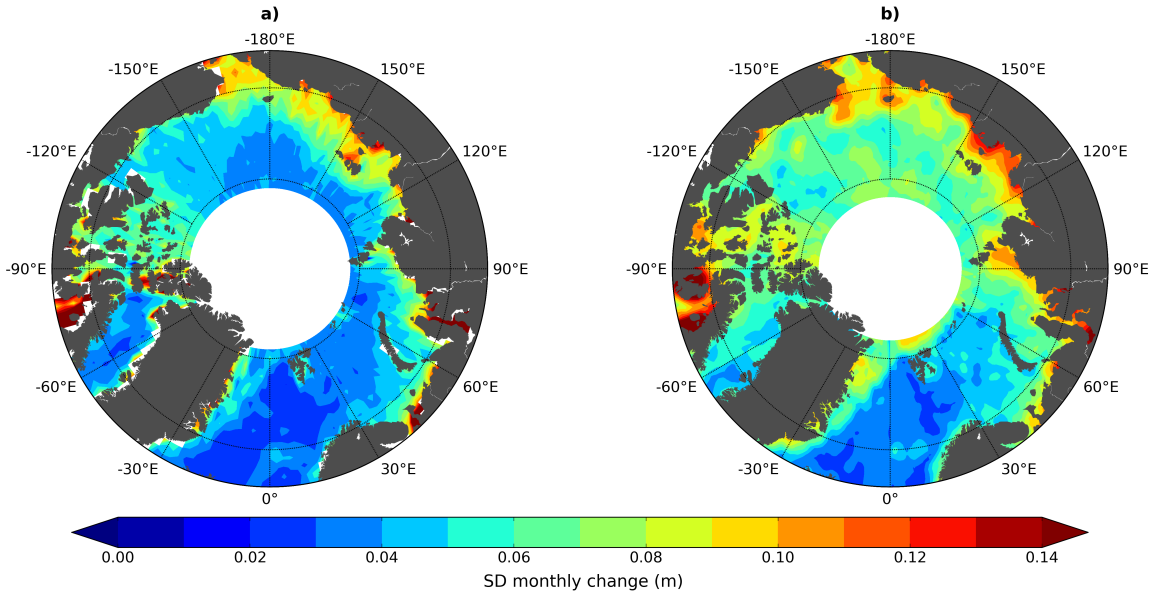

Figure 17: Standard deviation of monthly change in the Arctic sea level anomaly from June 2002 to May 2010: a) computed for the new solution, b) computed with DTU Arctic sea level anomaly maps.

Response to Reviewers' Comments

Manuscript TGRS-2017-00491
(December 2017)

We would like to thank our reviewer for his analysis and his comments. We have revised our paper based on the new comments and add additional analyses and figures from its suggestions. We hope the revised paper is complete and clear enough. The answers to the comments are given below.

1) **This is the main issue I think you need to address: you need to do more to demonstrate the ice-covered-to-ocean SSH continuity aspect of the paper, given that it is such a central part of your paper (it's in the title). In the current draft the only evidence you have provided, besides some sound theoretical arguments, is 'eye-balling' a single monthly SLA map, which isn't enough. Small mean differences (<5cm) would be difficult to detect on a colorscale spanning >30cm, and using such large grid boxes (of order 100km) any bias between lead and open ocean SSH would likely be smeared out. One simple way to examine this is to find the difference between 'open ocean' SLA and 'lead' SLA in individual grid cells (maybe slightly smaller grid cells) for every month, and find the monthly mean. For you to claim that you are producing consistent SSH estimates across the ice edge to be true (I'm not saying they aren't!) I would expect the mean difference to be close to zero, and definitely within the uncertainty/spread of the monthly differences. This would also allow you to look at the lead-open ocean difference seasonally, which would help to address another of my points:**

   Can the authors comment on the effects of melt ponds on their data? I would expect that in early summer, a lot of the 'lead' waveforms will in fact be melt pond waveforms. This could lead to a seasonal bias in the 'lead' elevation, as ponds will not necessarily be at sea level, particularly early in the melt season (see Armitage et al. (2016)).

   We agree that discriminating melt pond waveforms from lead waveforms is challenging. At this time we do not have a clear solution. In fig. 5 we can see that the lead fraction identified by the neural network classification strongly increases in early summer. It is possible that a part of this increase is due to melt ponds. We have added a sentence on this point when discussing fig. 5.

   Indeed, in Armitage et al. (2016) we found that lead elevations were ~1-4cm higher than open ocean elevations in summer due to melt ponds (earlier in the summer) and re-freezing leads (in late summer/early autumn). Another option would be to examine along track SLA across the ice edge, however this would necessarily be limited in space and time. Either way, I think more needs to be done here.

We understand that this part needed more to demonstrate the SSH continuity between open ocean and the ice-covered ocean. We have tried to perform the statistical analysis of the SLA

difference between 'open ocean' and 'leads' in small grid cells (25km x 25km) but our conservative hybrid classification does not allow to find a statistically sufficient number of points in such small box. As a result, each box at the edge only contains a small number of points which is not statistically relevant. As we understand that we must demonstrate the continuity at the very ice edge, we choose to show along-track diagnoses as you also suggested. We have thus added 3 new diagnoses: Fig. 15 a, b and c representing along-track SLA at the ice edge for one particular track (pass 788) but for 3 cycles: 64 (December 2007), 66 (March 2008) and 69 (June 2008, during the melting period). We think this plot shows the data and our results in a more detailed way and allows to present the processing quality. We can see that we also have the melt pond issue you mentioned, illustrated on cycle 69 and causing an increasing of lead SLA of few centimeters. We have described this figure and this behavior in a new paragraph in chapter 5.2. We hope the continuity part is more clear and complete now.

2) **page 2, line 28, a reference to Armitage et al. (2016) is missing here.**

The reference has been added here.

3) **Page 14, line 10-13: A paper relevant for discussions of snagging is Armitage & Davidson (2014) where we showed using the CryoSat interferometer that even peaky 'lead' waveforms can originate from up to 1500m away from the nadir point.**

We have added a line including this reference, it was an oversight.

4) **Discrimination/Retracker 'accuracy'**

Section 3.1: I appreciate your summary of the potential weaknesses of just using a threshold retracker. It is worth noting, I believe, that as you note this kind of retracker is generally adapted because it is robust, and because it consistently picks the same point on the leading edge of the waveform. You should include a comparison of your new retracking technique against traditional threshold/Gaussian fit retracking techniques. Make a map of the mean difference and RMS difference for the lead waveforms, and maybe even a time series. Only then can you quantitatively say what potential improvement your new retracking offers. The difference is only really important if it varies significantly in space/time, otherwise it can be accounted for with an empirical correction. Note that the sea level comparison later in the paper includes differences due to all of the cumulated processing steps, and you can't isolate the contribution of retracking, or waveform discrimination.

We agree with the reviewer that, threshold retrackers are robust and are widely used specifically because of this feature. However, picking the same

point on the leading edge doesn't mean that it corresponds to the surface level. Depending on the waveform sampling and the surface roughness, the tracking point can move in the leading edge from the middle towards the echo maxima.

In this paper, we want to propose a new retracker that allows one to account for the effects of the surface roughness and the waveform sampling for peaky echoes in LRM, in a similar way as Kurtz et al. did for SARM (Kurtz et al. (2014)). The idea is to show our new solution but not necessarily saying that threshold retrackers give poor estimates. Of course, there will be differences and geographical patterns will appear on the map of the mean differences but it will be challenging to explain what are the causes. We have an illustration of the difference in retrackers for AltiKa (rather than ENVISAT). Here we have compared two approaches for the sea- ice freeboard — one with the Enhanced Brown (adaptive) retracker and the other the standard retracker, Ice1. As the tracking over floes is essentially the same, this shows the difference in determination of the sea level in the leads. We clearly see significant geographical differences:

but determining the which solution gives the right estimation everywhere requires extra analysis if we want to be perfectly rigorous, and goes beyond the objectives of the present study.

I agree with your arguments, both here and in the manuscript. However, you haven't shown which techniques give the most accurate retracking/discrimination, especially given the now fairly large array of techniques in the literature for both of these processing steps. You have presented the theoretical arguments of why you might expect this (which, as I say, I broadly agree with, particularly for the retracking), but you haven't provided evidence that it is true generally against all other possible techniques. Thus, I don't think it is reasonable for you to claim, particularly in the abstract (page 1, line 18), that your methods give: "improved detection of water surfaces" or "more accurate water levels". The simple answer is to remove the words "improved" and "more accurate", and the sentence still makes perfect sense.

Of course, it is impossible to bring the absolute proof of which solution is the reality (or the closer to the reality) without a comparison with in situ data. That is why we have removed the two words you suggested. Nevertheless, we have added SLA estimates computed using the standard ocean retracker and the ice1 retracker to the Fig. 15 respectively on measurements classified as 'open ocean' and 'leads'. In this way, we want to show that using our retracking solution, the same algorithm to process both open ocean and lead waveforms, we have a better consistency than combining 2 different algorithms such as an ocean retracker and a threshold retracker (Ice1). And we can see that the bias between the ocean retracker and the Ice1 retracker is not a constant but evolves with time. We know that it is not the absolute

proof, but we think this comparison speaks for itself and illustrates the theoretical arguments exposed upper in the paper.

5) **Page 16, lines 40-44: we addressed seasonal SSH variability due to the seasonal input of freshwater to the Arctic in Armitage et al. (2016) – note that seasonal warming has a negligible effect compared to seasonal freshening since the salinity expansion coefficient is much (around 30 times(?)) larger than the thermal expansion coefficient at Arctic ocean temperatures. Similarly, we examined mecoscale variability using altimetry (eddy kinetic energy) in Armitage et al. (2017).**

We have added the 2 references (Armitage et al. 2016 and Armitage et al. 2017) in this part, thank you for the precisions, it is very interesting.