This document contains the pointers where revisions addressing reviewer comments can be found. This is accompanied by the revised manuscript with changes highlighted built using latexdiff.

Reviewer 1:

Major comment 1, implications to thickness retrieval: Section 4, second to last paragraph.

Major comment 2, Section 3 hard to follow: Whole Section 3.

Open water / lead / open ocean: Throughout the text.

Show the parameters for the two test periods in the same plot: Figures 3 and 4. Note that we chose to show two 15 day periods in the same month in one plot to give an idea about the change within weeks.

P4121 L9 The (pulse-limited) across track footprint: Subsection 2.1, last paragraph

P4121 L26, spillover: Addressed already in the comment.

P4123 L20 and P4124 L12, 128/256 bins: Subsection 2.1, first paragraph.

P4125 L7ff The authors should add a short description of the basic concept of a k-NN classifier: Subsection 3.2.

P4126 L1ff The description in this part is difficult to follow: Whole Section 3

P4126 L6 "Ties are broken at random": Subsection 3.2. Third paragraph.

P4126 L13ff What are the actual limits for the scaling: Subsection 3.3. Fourth paragraph

P4126 L14 But how is the distance between feature vectors used in the classification scheme? Really whole section 3, especially 3.2 and 3.3.

P4127 L2 Suggestion: Describe TPP as late tail to peak power and KF1 as early tail to peak power: Throughout the text, especially Eq 3 and 4.

P4127 L23 Move the last sentence to the next paragraph. It reads as if the LEW > 14 is used for the removal of leads: Subsection 3.3, paragraph 10

P4128 L12 Is there a reason to smooth the feature space and not the waveforms? Already addressed in comments

P4128 L20 The authors should add a description how the classification is done: Subsections 3.2 and 3.3.

P4129 L11ff If the classifier takes the next 3-5 neighbors...: Subsection 3.2, last paragraph

P4130 L6ff Also the incidence angles..: Already addressed in comments.

P4131 L6 Add "following" before "five days test set": Subsection 3.3, paragraph 3.

P4132 L26ff For the very thin ice it is crucial to define the term open water: Now “open ocean” is used for the open ocean throughout the text, and open water thus should be understood by the reader simply as non-ice covered water.

Can the authors check whether there are a higher levels of lead detections in the areas

which are labeled as thin ice? Already in the comments.

Also in November the young FYI has a significant misclassification as MYI. Would this

be in areas that were ice covered by the end of the melting season? Already addressed in the comments.

P4133 L24ff Would NRT ice thickness information be at least helpful for ice charting? Already addressed in the comments.

Table 1 2 Add a description of the rows and columns (columns: AARI classifications?): Captions of the tables 1 and 2.

Reviewer 2:

Fig 1 and 2: This we did not change, rationale given in the response.

Eqn 1 and throughout (128/250 bins): Subsection 2.1, first paragraph.

P 4124, L4-5: The SSD is essentially: Subsection 3.1, third paragraph.

4124, L 20-23: This is a vague sentence: Subsection 3.1, between equations 3 and 4.

P 4125 L 13: During which time of year was the training data set initialized? Subsection 3.3, second and third paragraphs. The second mentions the 15 day period and the third states that the test data always comes immediately after the 15 day training period.

Andrew Shepherd:

UCL NRT Product validation: Section 5, last paragraph.

Manuscript prepared for The Cryosphere Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 28 November 2015

Utilisation of CryoSat-2 SAR altimeter in operational ice charting

E. Rinne and M. Similä

Finnish Meteorological Institute, Marine Research, P.O. Box 503, 00101 Helsinki, Finland

Correspondence to: E. Rinne (eero.rinne@fmi.fi)

Abstract

We present methods to utilise Cryosat-2 (CS-2) Synthetic Aperture (SAR) mode data in operational ice charting. We compare CS-2 data qualitatively to Synthetic Aperture Radar (SAR) mosaics over Barents and Kara seas. Furthermore, we compare the CS-2 to archived operational ice charts. We present distributions of four CS-2 waveform parameters for different ice types as presented in the ice charts. We go on to present an automatic classification method for CS-2 data which, after training with operational ice charts, is capable of determining open ~~water~~ ocean from ice with a hit rate of $> 90\,\%$. The training data is dynamically updated every five days using the most recent 15 days CS-2 data and operative ice charts. This helps the adaption of the classifier to the evolving ice/snow conditions throughout winter. The classifier is also capable of detecting three different ice classes (thin and thick first year ice as well as old ice) with success rates good enough for the output to be usable to support operational ice charting. Finally, we present a near real time CS-2 product just plotting the waveform characteristics and conclude that even such a simple product is usable for some of the needs of ice charting.

1 Introduction

Our aim is to present new methods to utilise satellite altimeter measurements in operational ice charting. We present an automatic classification method to derive different ice stages of development from Cryosat-2 (CS-2) waveforms. This is different from the most common sea ice application of satellite altimeters today, which is measuring the freeboard and the thickness of Arctic winter sea ice, see for example (Laxon et al., 2013).

The use of altimeters to support ice mapping has been suggested already 35 years ago by Dwyer and Godin (1980). Ice detected utilising altimeter waveforms has been compared with sea ice extents from passive microwave satellite instruments (Laxon, 1990) and ERS-1 altimeter based sea ice estimates were faxed to research vessels navigating in the Southern seas in the early 90's (Laxon, 1994). Mostly due to Synthetic Aperture Radars

(SAR) becoming the standard tool in operational sea ice charting, altimetry has developed into a method for climate research (Laxon et al., 2003, 2013; Giles et al., 2008; Kwok et al., 2009). Altimetry is, however, widely used in numerical weather prediction where fast delivery products from different altimeters on open ocean are assimilated into weather models (Vidard et al., 2009).

The diminishing Arctic sea ice cover is opening new sea routes. In consequence, navigation in seasonally ice covered waters is due to increase rapidly. This calls for accurate and timely sea ice information, especially on dangerous sea ice conditions. Vessels navigating in or near sea ice can be roughly divided into those wanting to completely avoid ice and those that can safely operate in medium first year ice. For the ice avoiding ships, ice edge detection is enough but for the latter group some information on the stage of development of the ice is needed as well. Most widely used Earth Observation (EO) data in operational ice charting are Synthetic Aperture Radar (SAR) frames. In areas of heavy traffic, such as the Baltic Sea for example, SAR data is virtually indispensable. However, because the number of SAR acquisitions is limited, other instruments, such as altimeters, may provide valuable additional information on sea ice.

Due to the number of SAR satellites flying today being reasonably small, ice services can face incidents when fresh SAR data is simply not available. The loss of Envisat and its ASAR instrument in April 2012 pointed out how dependent European ice charting was on a single satellite. Furthermore, SAR images have a limited spatial coverage which results in data gaps. For ice covered seas with little or only sporadic traffic, SAR frames are often not acquired. This is the case with Southern ice covered seas – as the number of ground stations in the Antarctic is small, SAR acquisitions over Southern sea ice would use costly satellite mass storage. This is different in the Arctic seas: for example the SAR frames from Kara and Barents seas could be directly downlinked to a ground station in northern Europe. Thus it is convenient to study altimeter ice charting in the European sector of the Arctic even if one of our aims is to contribute to Antarctic sea ice charting.

In absence of SAR data, ice services look for auxiliary data. Optical satellite images, such as MODIS or Suomi-NPP images may mitigate the problem, but only if the lighting

and cloud conditions are favourable. Alas the polar night and frequent cloud cover often render optical imagery useless. In these cases ice services are bound to either use low resolution (at the best in 10 km grid) products from, for example the EUMETSAT OSI-SAF project (http://saf.met.no/p/ice/) or, in the worst case, inform the users that up to date ice information is not available.

Zygmuntowska (2014) examined the possibility to distinguish between first year ~~and multi year~~ (FY) and multi-year (MY) ice throughout the winter using altimeter waveform parameters. They gridded their data in a 25 km grid and studied the parameters one at a time and found that the spatial distribution of a single parameter varied strongly from month to month from the freezing-up period to the Arctic spring. Sometimes the patterns followed the MY/FY ice areas, other times not. This we can regard as manifestation of sensitivity of the altimeter waveform for the ice/snow surface conditions.

2 Data

Our study area is the Barents and Kara seas in the European sector of the Arctic Mediterranean. We chose the area because of the good availability of data, namely the Arctic and Antarctic Research Institute (AARI) operational ice charts and an archive of SAR data as a heritage from the Enhanced Arctic Sea Ice Information – ANISTIAMO exercise carried out by FMI in 2014 (ANISTIAMO Reports, 2014) and its predecessor described in Similä et al. (2013).

2.1 Cryosat-2 data

We use the CS-2 SAR mode (Wingham et al., 2006) Level 1b (L1b) data (Bouzinac, 2014) available online from ESA. For this study we use the Near Real Time CS-2 products that were made available by ESA for our study. This product was built using the Baseline-B CS-2 processor, and thus differs from the current Baseline-C CS-2 product. Most importantly, the full range window in the Baseline-B product is 128 range bins whereas in Baseline-C it is

256 range bins (Scagliola, 2014) . This should be taken into account if our methodology is to be applied on the current Baseline-C products.

The main instrument of CS-2 is the Synthetic aperture radar/Interferometric Radar Altimeter (SIRAL) altimeter. The CS-2 L1b product used here is, essentially, the CS-2 SAR-processed average waveform for each point along the ground track of the satellite. In the SAR mode, SIRAL employs the along-track beam formation (i.e. SAR processing) to generate a resolution cell of approximately 300 m by ~~1~~1.65 km ~~(Wingham et al., 2006)~~ (Scagliola, 2014) . The SAR mode of CS-2 is originally designed for ice covered seas (Wingham et al., 2006) . With the along-track resolution enhancement due to SAR processing enables smaller leads to be detected within the sea ice pack as was possible with CS-2's predecessor, the Envisat RA-2 (Laxon et al., 2013). Thus the CS-2 is operated in SAR mode over the sea ice covered seas. This combined with the polar orbit of CS-2 with an inclination of 92° results in a good coverage of CS-2 SAR mode measurements over the ice covered Arctic oceans.

2.2 Ice charts

We take the ice stage of development from the weekly ice charts published by the Arctic and Antarctic Research Institute (AARI) (Bushuev and Loshchilov, 2007). The AARI ice charts are available online from http://www.aari.ru/. We downloaded the charts from the current AARI website predecessor, an AARI FTP-server, in SIGRID3 format (SIGRID3 Manual, 2014). The ice charts for the Kara and Barents Seas provide estimates of total ice concentration (CT), the partial concentrations (CA, CB and CC), and the stage of ice development (SA, SB and SC) for the three thickest ice types for polygonal areas. To quantitatively study the effect of sea ice stage of development on altimeter waveforms, we rasterised the AARI ice maps into 2 km grids in Lambert equal area projection. We sampled these grids so that for each CS-2 measurement falling within the test area the ice characteristics for the ice chart polygons were fetched from the temporally closest AARI map. As the AARI maps are generated weekly, the largest time difference between the the different data sets is three days.

The ice stage of development in the AARI ice charts follows the World Meteorological Organisation (WMO) sea ice nomenclature. The WMO nomenclature defines seven different stages of development: nilas, grey ice, gray-white ice, thin first year (FY) ice, medium FY ice, thick FY ice and old ice. Three first stages are thin ice, which are defined as ice that does not form pressure ridges. The three FY ice categories are ice which is thick enough to ridge and has not experienced a whole melt season. The last stage, old ice, is ice that has survived at least one whole summer melt season. In the cryospheric community, old ice is often referred to as multi year (MY) ice. In this study we use only three different stages of development: thin (<70 cm) FY (WMO categories nilas, gray, gray-white and thin FY), thick (>70 cm) FY (WMO categories medium and thick FY) and MY ice (WMO category old ice).

The AARI ice charts are based on SAR and optical satellite images as well as reports from coastal stations and ships. The segmentation of images and subsequent interpretation and mapping of ice conditions are carried out by ice experts. The main purpose of the weekly ice chart is to show the spatial distribution and characteristics of sea ice. We have not found English language scientific publications discussing the accuracy of the AARI weekly ice charts. However, they are the only source of ice information for this area with adequate spatial coverage and temporal resolution for our study. Furthermore they are completely independent of CS-2 data, which makes comparing the two meaningful.

The area of the polygons in the AARI ice chart varies largely. In some cases the smallest diameter of the polygon is around 10–20 km, in some cases the largest diameter of a polygon is up to 100–200 km. Even if the AARI charts provides partial concentrations for up to three ice types for each polygon, we assign just one ice type to a single polygon. In the Sect. ~~3.1.~~ 3.3 we need to apply threshold for the partial concentration of the dominant ice class. For the automatic classifier discussed in Sects. 3.3 and 4.2, this threshold is 75 %. When we assign an ice class to a polygon in the test phase, no threshold is used and the partial concentration of the dominant ice type is often below 75 %. Thus a large fraction of the polygon represents in reality some other ice development stage than what we have labelled it to be. In consequence, an inherent inaccuracy is present in our reference data.

2.3 SAR composites

To visualise CS-2 waveform characteristics, we use synthetic aperture radar composites compiled from Radarsat-2 frames. These composites were originally compiled to be used as input data for a multisensor ice thickness chart for the Barents and Kara seas in the ANISTIAMO demonstration project in spring 2014, see www.arcice.org. The SAR data in Figs. 1 and 2 is used for visualisation only and is included because it is the most widely used EO data in operational ice charting and we assume most of our readers to be familiar with it.

3 Methods

We set out to build a classifier to retrieve the ice stage of development using only CS-2 waveforms and past AARI ice charts as input. In this section we introduce all the waveform characteristics which we have utilized in our analysis (Section 3.1). Then we review the classifier methodology in the Section 3.2. Finally we discuss the actual classification. We also describe the preprocessing and postprocessing steps of our classification procedure in Section 3.3.

3.1 Waveform statistics

We chose four characteristics, pulse peakiness (PP), leading edge width (LEW), late tail to peak power ratio (~~TPP~~LTPP) and stack standard deviation (SSD) to describe the CS-2 waveform. The PP, ~~TPP~~ LTPP and LEW are easily derived from the waveform and the SSD is delivered in the L1B data product (Bouzinac, 2014).

The PP is defined as:

$$\mathrm{PP} = 128 \cdot \frac{\sum_{i=1}^{128} P_i}{P_{\max}} \tag{1}$$

where P_i is the power in the ith range bin and P_{max} is the maximum power in one range bin in the waveform.

PP has been used previously to distinguish leads from ice floes in pulse limited radar altimeter data (Laxon et al., 2003). Classifying waveforms with high PP as leads is an integral step in traditional radar altimeter sea ice freeboard processing.

The SSD is taken from the CS-2 L1b data product. The SSD is essentially the standard deviation of ~~a set of Doppler echoes~~ power values from a common surface ~~location prior to stacking~~formed from a set of Doppler waveforms over different incidence angles (Wingham et al., 2006) . SSD has ~~bee~~ been used, in conjunction with PP, in the lead detection, e.g. by Laxon et al. (2013); Ricker et al. (2014) and Kurtz et al. (2014).

For the LEW, we use the difference between the bins retracked with $\rho = 10\,\%$ and $\rho = 90\,\%$ percentage ~~.~~ using an Offset Center of Gravity (OCOG) ~~retrackers~~retracker. The OCOG retracker returns the bin number where the received power count rises for the first time over the threshold value τ of:

$$\tau(\rho) = \frac{\rho}{100} \cdot \sqrt{\frac{\sum_{i=1}^{128} P_i^4}{\sum_{i=1}^{128} P_i^2}} \tag{2}$$

There the ρ is the percentage.

The OCOG retrackers have been widely used in altimetry, see for example Wingham et al. (1986), Bamber (1994) or Soussi and Femenias (2006).

~~The TPP~~ In our classification experiments we chose to experiment with the two features suggested by Kurtz et al. (2014) .The ratio of late tail to peak power is defined as:

$$\text{TPP}\text{LTPP} = \frac{\frac{1}{21}\sum_{i=\text{max}+50}^{\text{max}+70} P_i}{P_{\text{max}}} \tag{3}$$

where max is the index of the range bin with the maximum power.

~~TPP~~ LTPP tells us how much off-nadir power there is present in the tail of the waveform. Typically ~~TPP~~ LTPP is high for surface with large roughness.

We also studied the feasibility of another waveform charasteristic suggested by Kurtz et al. (2014), but found that TPP yield better results in the automatic classification. The rationale of choosing is discussed in Sect. 4.2. This is the ratio of early tail to peak power (denoted here ETPP):

$$\mathrm{ETPP} = \frac{\frac{1}{6}\sum_{i=\mathrm{max}+1}^{\mathrm{max}+6} P_i}{P_{\mathrm{max}}}, \tag{4}$$

The ETPP describes how rapidly the power P declines immediately after the maximum value.

On the detection of potential leads we employ three slightly different statistics to characterize specular reflections. These three features consist of PP and two statistics presented in Ricker et al. (2014) . They are are defined as follows:

$$\mathrm{PP}_{\mathrm{left}} = 9 \cdot \frac{P_{\mathrm{max}}}{\sum_{i=\mathrm{max}-1}^{\mathrm{max}-3} P_i} \tag{5}$$

and

$$\mathrm{PP}_{\mathrm{right}} = 9 \cdot \frac{P_{\mathrm{max}}}{\sum_{i=\mathrm{max}+1}^{\mathrm{max}+3} P_i} \tag{6}$$

where max refers to the index of the range bin with maximum power. We determined experimentally the thresholds: PP > 40 and either $\mathrm{PP}_{\mathrm{left}} > 20$ or $\mathrm{PP}_{\mathrm{right}} > 15$. Ricker et al. (2014) also used a SSD threshold for lead detection. However, in our data set when the PP conditions were met, SSD was almost always below 4 and we did not set an additional condition.

It is important to note here that we rely only on the four waveform characteristics: PP, SSD, TPP and LEWwaveform characteristics. We do not do freeboard processing in the style of

Laxon et al. (2013) or use the freeboard values in the higher level CS-2 data products. This is because we want the CS-2 methods to be simple as well as independent from any other data sources.

3.2 ~~Automatic~~ k-Nearest Neighbors classifier

We ~~set out to build a classifier to retrieve the ice stage of development using only CS-2 waveforms and past AARI icecharts as input.~~ review shortly the basic ideas of the k-Nearest Neighbors (k-NN) classifier. The k-NN classifier is a simple, memory based classifier which does not require a statistical model for the data. The k-NN classifier can in some cases achieve the error rates similar to the Bayesian classifier which is the statistically best classifier (Hastie et al., 2001) but a significantly more complicated one than the k-NN.

At the first phase we collect the training data together into a set of feature vectors and corresponding classes. In our case the features were the waveform based statistics (SSD, LEW, PP and LTPP) and the classes are the ice stages of development (open ocean, thin FY, thick FY and MY ice).

To decide the class of a new sample point from a test set, we find out the k closest samples from the training data and perform a majority vote among them. The mode class among the k closest neighbors is selected as the class for the new sample point. If two or more classes have the same amount of samples in the group of k closest neighbors, then the class is selected randomly from those classes.

The crucial requirements for the efficiency of the classifier are:

1. The training set must represent well the data to be classified.
2. Dimension of the feature space can not be very large.
3. The number k of the neighbors must be determined soundly.
4. The distance between the measurements must be a proper metric.

The class boundaries in k-NN are determined totally by the data. Hence, the first requirement is of major importance. The second requirement is natural for all classifiers.

One manifestation of the "curse of dimensionality" is that in high dimensions, almost all pairs of points are equally far away from each other (see Hastie et al. (2001)). The suitable amount of the neighbours k depends on the dimension of the feature vector and the size of the training data. Here k is determined empirically. We used the ordinary Euclidean metric as our distance function.

It is worth to emphasize that the distance between the sample points is measured in the feature space. Two neighboring points in feature space can be spatially hundreds of kilometers away from each other. Closeness in feature space only implies the similarity of waveforms.

3.3 Classification procedure

Due to the reasons discussed below we selected a k-nearest neighbours (k-NN) classifier for our classification method. The adopted classifier is able to classify four different ice classes (open ~~water~~ocean, thin FY, thick FY and MY ice) with a reasonable accuracy.

The results of Zygmuntowska (2014) lead us to adopt an approach where we dynamically update the training set for the classifier. The training data was gathered from the CS-2 acquisitions and the AARI charts during a 15 days period. ~~Then~~ Only the AARI polygons where the partial concentration of the dominant ice class was $> 75\,\%$ were used for training. The data was divided into four different ice classes. Using the k-NN classifier we then determined the class boundaries for the training data set.

Finally we used these class boundaries ~~in analysing~~ to process the CS-2 data of the following five ~~days~~ day period, called here test data. The training data~~and, hence~~, and hence also the class boundaries, were recalculated at intervals of five days. We chose to do this, instead of applying fixed class boundaries all the time~~. The~~ , because the constantly updated class boundaries help the classifier to adapt to ~~ice/snow conditionswhich evolve continuously~~continuously evolving ice and snow conditions. The magnitude of the change in the training data can be seen in Figs. 3 and 4 as the difference between the dashed and solid lines. The difference is subtle, but according to our study, large enough to cause problems if not taken into account.

As one can see from Figs. 3 and 4 the waveform parameters of different ice categories are often very close each other. In our data set this is true especially in November during the freeze-up period. It is obvious that we cannot separate the ice types using just one parameter. Instead a set of parameters should be utilised simultaneously. The shape of distribution of a single parameter for ice class deviated in most cases essentially from the normal distribution. The task to non-parametrically model a non-Gaussian multidimensional distribution is very challenging. Hence, we chose not to use a Bayesian classifier which is the theoretically optimal classification method (Hastie et al., 2001) .

We avoided the difficulty to deal with non-Gaussian multidimensional model the multidimensional non-Gaussian distributions by applying the k-nearest neighbors classifier. It's largest advantage is that it does not require a statistical model for the data. The classifier is memory based. The concept of the classifieris simple. We first collect all our training data together. When the class of a new feature vector from a test set is decided, we first search k closest training points and perform a majority vote among them. Ties are broken at random. The crucial requirements for the efficiency of the classifier are: the training set must represent well the data to be classified, dimension of the feature space can not be very large, the number k of the neighbors must be determined soundly and -NN classifier, see Section 3.2.

In order to calculate the distance between the measurements must be a proper metric.

We first discuss the distance metricfeature vectors we used the Euclidean distance with equal weights. We scaled the distribution of each parameter on the interval $[0, 2]$. The full range interval $[0, 2]$ corresponds the PP value range $[0, 40]$, LEW range $[0, 8]$, SSD range $[0, 50]$ and finally LTTP range $[0, 0.18]$. The ranges were determined empirically. The values of rare events were truncated. Naturally, we could have used also the unit interval $[0, 1]$. Now the distance between feature vectors can be computed using the ordinary Euclidean distance.

During the 15 days training period we collect a rather extensive data set. In most cases this set covers the variation occurring in the following five days CS-2 data. To characterize the waveform we selected first the standard features (PP, LEW, SSD) as a part of a feature

vector. Increasing the dimension of the feature space always ~~reduces the amount of training~~ increases the mean distance between sample points inside a unit cube ~~. In consequence, a situation where a new feature vector does not have any training point close to it begins to occur more frequentcly andthe classification accuracy decreases~~as noted in Section 3.2. This implies that even small changes in the training data and, hence, in the class boundaries may have a significant impact on the classification result.

On the other hand, the discrimination between feature vectors is, in principle, more efficient if the description of a waveform is more versatile, i.e. the number of features is larger. Keeping in mind these these two ~~conflicting~~ consequences of adding features,we studied if an additional feature would ~~indeed improve~~ improve or weaken the results. We chose to ~~experiment with the two features suggested by Kurtz et al. (2014) . The first one is the ratio of tail to peak power (denoted here Kf1):~~

~~$$\mathrm{Kf1} = \frac{\frac{1}{6}\sum_{i=\max+1}^{\max+6} P_i}{P_{\max}},$$~~

~~This feature describes how rapidly the power $P_{\max}$ declines immediately after the maximum value. The other feature proposed was the ratio of waveform tail to peak power or TPP as defined in Eq. (3) in Sect. 3~~test the influence of adding the LTPP and/or ETPP characteristics.

A way to measure dependence between two random variables is to utilize the mutual information (MI) (Cover and Thomas, 1991). MI measures the overall dependency between two random variables, not just the linear dependency like the correlation coefficient. We calculated the pairwise values of MI between features ~~Kf1 and TPP~~ ETPP and LTPP and the set of standard parameters (PP, LEW, and SSD). The MI values were consistently lower for ~~TPP than Kf1~~LTPP than ETPP. This implies that the ~~TPP~~ LTPP contains less overlapping information with the standard three parameters than ~~Kf1~~ETPP.

Next we checked if the addition of ~~TPP~~ LTPP to the three previous parameters (PP, LEW and SSD) actually improves the k-NN classification. The result was that the addition of ~~TPP~~ LTPP increased the correct classification accuracy from 0 to 5 % for a single class depending on the ice class and the test set. This was more than the improvement after addition of

~~Kf1~~ETPP. We also tested a 5-dimensional (5-D) feature vector containing the parameters (~~Kf1, TPP~~ETPP, LTPP, PP, LEW, SSD). The addition of ~~Kf1~~ ETPP had a negative impact on the accuracy compared to the tested 4-dimensional (4-D) feature vector. Hence, we selected the features (PP, LEW, SSD, ~~TPP~~LTPP) as our feature vector.

Prior to the classification we preprocess the data. All waveforms with LEW larger than 14 are excluded from analysis. We regard these waveforms too noisy to be useful. ~~We also want~~

We also wanted to remove potential leads from data to limit the confusion between different ice types.

~~On the detection of potential leads we employ three slightly different statistics to characterize specular reflections. These three features consist of PP and two statistics presented in Ricker et al. (2014) . They are are defined as follows:~~

$$\mathrm{PP}_{\mathrm{left}} = 9 \cdot \frac{P_{\mathrm{max}}}{\sum_{i=\mathrm{max}-1}^{\mathrm{max}-3} P_i}$$

~~and~~

$$\mathrm{PP}_{\mathrm{right}} = 9 \cdot \frac{P_{\mathrm{max}}}{\sum_{i=\mathrm{max}+1}^{\mathrm{max}+3} P_i}$$

~~where max refers to the index of the range bin with maximum power. We determined experimentally the thresholds:~~ This was done by removing all waveforms with $\mathrm{PP} > 40$ and either $\mathrm{PP}_{\mathrm{left}} > 20$ or $\mathrm{PP}_{\mathrm{right}} > 15$ ~~. Ricker et al. (2014) also used a SSD threshold for lead detection. However, in our data set when the PP conditions were met, SSD was almost always below 4 and we did not set an additional condition~~(see subsection 3.1) as potential leads. The amount of ~~potential lead signals was usually 5–7~~lead detections was usually 5–13 % from the measurements~~. Waveforms considered potential leads were excluded from the analysis~~, depending on the season and ice type. The amount of potential leads was larger in November than in March. Also the number of potential leads was larger for thin ice class than for thicker ice types.

To reduce speckle, we average the CS-2 waveform parameters over the five consecutive footprints and assign the value to a single footprint area, i.e. we use the running mean method for each parameter separately. ~~All~~ The CS-2 measurements are subject to speckle like any other SAR signature. The speckle influence is most obvious for PP but it is also present in other waveform parameters. In this step we implicitly assume that the five consecutive footprints covering a track of 1900 m belong to the same ice class. Considering our coarse typing of ice classes (3–4 ice types) this is a reasonable assumption.

We classify the running means of the waveform parameters using the k-NN classifier. At the first phase we perform the classification for each 4-D feature vector separately. Then we take the mode of 50 consecutive ice class labels. The resulting mode is then regarded as the estimated ice type and it has a resolution of 19 km. If we take into account the spatial averaging the true resolution is about 20 km along the track. We examined also the possibility to use only 30 consecutive class labels to achieve a ~~better spatial resolution~~ spatial resolution of about 12 km. In our test runs the 20 km resolution yielded, however, slightly more accurate results than the 12 km resolution. Although the difference was not large we chose the coarser resolution data because of slightly better reliability.

The mode operation is also applied for our ground truth data points extracted from the AARI ice charts (50 consecutive points corresponding the CS-2 points). When we assess the classification accuracy using the ground truth data, it takes place comparing these 19–20 km long segment tracks. In the classification maps in the Sect. 4.2 we show the results using a sliding window technique, i.e. the consecutive the class labels have just a distance of 0.38 km between them.

We still must determine a reasonable value for k. That is, how many feature vectors from the training data do we use to build class boundaries. The value $k = 1$ yielded highly variable results for the test sets and was deemed unpractical for our purposes. When we compared the values $k = 3$ and $k = 5$ we noticed that the results were close to each other for ~~the Novembertest sets (3 sets)~~ November, although $k = 3$ yielded slightly better results. In March the results more clearly favored $k = 3$ over $k = 5$. Increasing k to a larger value than $k = 5$ lead to poorer results. The difference between the November and March data

sets was that the number of ice types was four in March and three in November. Hence the classification task in March was more challenging than in November. We have used the value $k = 3$ in our classifications.

~~The disadvantage of the k-NN classifier with the mode operation is that the classes with most measurements are usually best detected.~~

4 Results and discussion

4.1 Qualitative comparison

To visualise the behaviour of PP, Fig. 1 below shows PP drawn over a SAR composite from the same area on the 05 March 2014. The figure shows an increase from the low (PP < 3, blue) values over open ocean to high (PP > 7, red) values over ice. The increase coincides with the ice edge visible in the SAR frame. Analogously to the Fig. 1 and PP, Fig. 2 shows the CS-2 SSD over a SAR frame. Again the ice edge is clearly visible as SSD changes from high values (SSD > 20, blue) over open ~~water~~ ocean to small (SSD < 10, red) on ice. However, SSD seems to exhibit more variation over ice than PP does. There are areas of high SSD within the ice pack, some of which coincide with features in SAR data. The interpretation of a SAR frame to ice characteristics is subjective due to the multitude of factors affecting the backscattering. Areas of high backscatter can be thick heavily deformed ice or, in some cases, broken thin ice with varying amount of open ~~water~~ ocean e.g. brash ice. Detailed information on SAR backscattering statistics in our test region can be found in Lundhaug (2002). Due to the ambiguity in SAR signature, it is impossible to derive the ice thickness from a single SAR frame alone. Because of this we ~~do~~ did not carry out a quantitative comparison of SAR and CS-2 data.

To further study the effect of ice stage of development on the CS-2 waveform, we sampled the AARI ice charts at a point closest to the CS-2 measurement as described in Sect. 2.2. The distributions of waveform parameters for different stages of development are shown in Figs. 3 and 4. Open ~~water~~ocean, as expected from the SAR/CS-2 comparison above, shows

peaks in high (SSD > 50) SSD and low (PP < 5) PP. In March thin FY ice has a bimodal PP distibution, possibly due to the polygons labeled with thin ice often having a total ice concentration of less than $100\,\%$, ie. there are ~~areas of open water~~ open water areas present in addition to the ice. Overall, the distributions show promise for distinguishing different ice classes based on the four waveform parameters. ~~The results from out automatic classifier~~ Our automatic classifier results are presented in the next subsection.

4.2 Automatic classification

The classification methodology is presented in Sect. 3.3. The waveform parameters are correlated but have also differences in their distributions as shown in Figs. 3 and 4. We utilised three different ice categories in November (open ~~water~~ocean, thin FY, and MY ice) during the freeze-up period and four ice categories in the middle of winter in March (open ~~water~~ocean, thin FY, thick FY, and MY ice).

In the training data the dominant ice type from AARI charts was used as the true ice type for all CS-2 waveforms falling within the ice chart polygon if the partial concentration of the dominant ice type was $> 75\,\%$. Only measurements from this kind of polygons were accepted for training. Then the CS-2 data for the following five days test set (see Sect. 3.3) was classified using the system. The results of the CS-2 classification were then compared to the stages of development taken from the temporally closest AARI chart during the test period. In the test set we do not use the $75\,\%$ rule for the dominant ice class, but use the stage of development which has the highest concentration as the truth for all of the CS-2 measurements falling within the polygon. Tables 1 and 2 show the classification matrices for November and March, respectively. Maps of the classification results are presented in Figs. 5 and 6 for November and March, respectively.

In November thin FY ice, MY ice and open ~~water~~ ocean are present both in the CS-2 data as well as in the AARI chart (Fig. 5 and Table 1). Thick FY ice was absent in the AARI charts we used as training for November, and thus we only have three classes: open ~~water~~ocean, thin FY and MY ice. The open ~~water~~ ocean is classified right in $98\,\%$ of the cases. The thin FY ice mixes somewhat with MY ice: $46\,\%$ of CS-2 measurements in polygons marked

to consist mostly of thin FY ice in the AARI charts are classified to be MY ice based on CS-2. Analogously 8 % of CS-2 measurements from polygons where MY ice is dominant is classified as thin FY ice. Part of the inconsistecy is natural. In reality there are inclusions of FY ice within the old ice area as well as inclusions of MY ice in the FY ice area. However, there are CS-2 measurements classified as MY ice south of 80° N. It is unlikely that these are in reality MY ice. We assume these to be areas of deformed ice where the large scale surface roughness is more akin to MY ice than recently formed FY ice. If this is the case, the information about deformed ice, most likely an obstacle to navigation, would be valuable for operational ice charting. Sadly, we have no means to test our assumption.

For March (Fig. 6 and Table 2) the results are similar to November. The overall correspondence of AARI maps and the CS-2 classification is good. The two FY ice classes mix considerably. This is not an unsurprising result because the thickness of 70 cm is not a threshold which would abruptly change the characteristics of ice. Instead the division of ice thickness of less or larger than 70 cm is mostly based on the needs of ice navigation. Furthermore there are inclusions of thin FY ice within the thick FY and vice versa. The results for open ~~water~~ ocean (93 % right) and MY ice (83 % right) are good. A notable feature in the chart is MY ice appearing in the CS-2 measurements at about 78° N and 95° E, West of Vilkitsky Strait. There were small amounts of MY ice near the coast marked in the AARI charts too~~, especially near the coasts~~. Thus it may be that our classification exaggerates the amount of MY ice, especially in the areas where heavy deformation is likely to occur. However, for the purposes of operational ice charting, a cautious approach is often preferred.

Our classification results are similar to those of Zygmuntowska et al. (2013), obtained for an airborne altimeter in a smaller scale. Zygmuntowska et al. (2013) used a Bayesian classifier and presented a comparison of CS-2 derived sea ice types and OSI-SAF ice types. They showed that ice type classification with satellite altimeter data is possible but also found regions where clear discrepancies occur between the CS-2 derived ice type and their validation data. They attributed these discrepancies to areas of FY ice with large surface roughness. We also found false positive MY ice classifications. As discussed before,

for our application, classifying heavily deformed FY ice as MY ice is not a problem since both present similar threat to navigation.

Detection of MY ice has implications to sea ice thickness retrieval from CS-2. Most of the current altimeter sea ice thickness processors apply some kind of a MY ice mask, firstly to modify the snow climatology used (for example Laxon et al. (2013)) and then to modulate the ice density for freeboard to thickness conversion (Kern et al., 2014) . Often used source for the ice type is the OSI-SAF ice type product (for example Laxon et al. (2013) and Ricker et al. (2014)). Our methodology produces realistic results, especially in March, for detection of MY ice and thus could benefit the traditional ice thickness retrievals. However, here the false positive MY detections would be a larger problem than they are for the ice navigation application. The density of heavily deformed FY ice would still be higher than the density of MY ice and thus false MY detections would result in too small thickess. Furthermore, the MY mask could easily be derived from the operational ice charts directly, without any waveform based classification necessary. Having said that, our study does prove that the C-S2 waveforms contain information about the sea ice type. For near real time applications of CS-2 data MY ice mask would be an interesting option since no auxilliary MY products for the time of the measurement are yet available.

When we inspect the classification results, the detection of thin ice (here FY ice thinner than 70 cm) has been least successful. As seen in Figs. 3 and 4 the waveforms originating from thin ice exhibit a wide range of variation for all used features. This is understandable. Very thin ice (thickness less than 10 cm) has often ice concentration well below 100 %. Due to this, part of the waveforms assigned to thin ice are actually waveforms from open water. This can be seen especially clearly in the behaviour of SSD in Fig. 2. Processes such as rafting and ridging increase surface roughness, and thin ice as defined in this work can be easily ridged. The thin ice is mostly mixed with thick FY ice (thickness larger than 70 cm) but to a lesser degree also with open ~~water~~ocean.

5 Cryosat-2 product to support FMI operational ice charting

We built a CS-2 ice product based on near real time waveform characteristics to test if the Finnish ice service would benefit from CS-2 data. The product is basically maps of Near Real Time (NRT) CS-2 PP, LEW and SSD. A comparably simple system was built to download the data from an ESA server and to calculate and plot PP, LEW and SSD. All of the CS-2 data acquired during the past two days is used for each product. The prototype system has been running at the Sodankylä satellite receiving station since September 2014. The FMI CS-2 product is available online at http://ice.fmi.fi/Cryosat/. The ice experts were told to look for abrupt changes in the PP to detect the ice edge and to interpret areas of constant low PP as open ~~water~~ocean. The analysts were also briefed that in addition to open ~~water~~ocean, heavily deformed sea ice may result in low PP. We did not endeavour to make an automated ice edge detection since the ground tracks are rather sparse and we did not want to interpolate between them.

The product was received well. During the fall the Finnish ice service provided ice information to ships in the Kara Sea and the CS-2 product was used as an independent reality check for Sentinel-1 and COSMO-Skymed SAR frames. As expected, in an area where SAR data is readily available, such as the Kara Sea, altimeter products add little or no value to operational ice charting. However, several requests were made to build a similar product for the Antarctic ice covered oceans. Our plan is to continue providing the FMI CS-2 product and implement a similar product for the Sentinel-3A data in the future.

In April 2015, University College of London (UCL) published a near real time CS-2 sea ice thickness product available online. The UCL product is based on the CS-2 processing chain presented in Laxon et al. (2013). ~~At the moment of writing, the UCL product lacks validation or comparison to independent EO and~~ The volume estimates derived from the UCL NRT and the standard UCL product have been shown to agree within 0.5 %. However, as no user experiences nor comparisons to independent data in ship scale are available, it is hard to assess ~~its usability~~ the usability of the UCL product for navigation. This will surely change as the UCL product becomes more well known and user cases begin to form. Furthermore,

a comparison of fine resolution data (such as ship measurements) to CS-2 data such as the UCL NRT product, would be a natural continuation of our study presented in this paper.

6 Conclusions

For the first time, we have demonstrated the use of a SAR altimeter, namely the SIRAL-2 onboard CS-2, to support operational ice charting. We've presented a qualitative comparison of a SAR composite and CS-2 data. Furthermore, we've compared the CS-2 waveform characteristics to the stage of development of ice taken from ice charts. We've also presented an automatic classification system capable of detecting open ~~water~~ocean, thin FY ice, thick FY ice and MY ice based on four CS-2 waveform characteristics. This is the first time ice classification methodology using satellite SAR altimeter data is presented and tested outside gray literature. The classification system requires recent operational ice charts for the training, but after the training the only input are CS-2 data. The system resolves MY ice and open ~~water~~ ocean well. The two tested FY classes mix significantly, but for the application of operational ice charting this is not a problem.

We have built a prototype system providing simple maps of NRT CS-2 waveform characteristics. This product has been tested by the Finnish ice service during the winter of 2014–2015. The feedback was positive. Thus we conclude that satellite altimeters, in this case the CS-2, provide an independent source of sea ice information to complement SAR and passive microwave data. Albeit low resolution and sparse, altimeter measurements can be used at times and locations where other data sources are unavailable.

Acknowledgements. The authors wish to thank ESA for providing the NRT CS-2 data. We thank the people of the Finnish Ice Service for their valuable feedback on our CS-2 product. We thank the Russian AARI institute for making their ice charts available. Finally, thanks to Juha Karvonen for providing us the SAR mosaics used in this paper.

References

ANISTIAMO Reports: Enhanced Artic Sea Ice (ANISTIAMO) Reports, Tech. rep., ESA, available at: www.arcice.org, last access: 27 July 2015, 2014.

Bamber, J.: Ice-sheet altimeter processing scheme, Int. J. Remote Sens., 15, 925–938, 1994.

Bouzinac, C.: CryoSat Product Handbook, ESA User Manual, ESA, ESRIN, Italy, 2014.

Bushuev, A. and Loshchilov, V.: Ice chart composition at AARI, in: Remote Sensing of Sea Ice in the Northern Sea Route: Studies and Applications, vol. 4, edited by: Johannessen, O. M., Springer-Praxis, Berlin, 243–252, 2007.

Cover, T. and Thomas, J. A.: Elements of Information Theory, John Wiley and Sons, New Jersey, USA, 1991.

Dwyer, R. and Godin, R.: Determining Sea-Ice Boundaries and Ice Roughness Using GEOS-3 Altimeter Data, NASA Contractor Report, Wallops Flight Center, USA, 1980.

Giles, K. A., Laxon, S. W., and Ridout, A. L.: Circumpolar thinning of Arctic sea ice following the 2007 record ice extent minimum, Geophys. Res. Lett., 35, L22502, doi:10.1029/2008GL035710, 2008.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning; Data Mining, Inference, and Prediction, Springer, New York, USA, 2001.

Kern, S., Khvorostovsky, K., Skourup, H., Rinne ,E., Parsakhoo, Z. S., Djepa, V., Wadhams, P., and Sandven, S.:The impact of snow depth, snow density and ice density on sea ice thickness retrieval from satellite radar altimetry: results from the ESA-CCI Sea Ice ECV Project Round Robin Exercise, The Cryosphere,9, 37–52,doi:10.5194/tc-9-37-2015, 2015.

Kurtz, N. T., Galin, N., and Studinger, M.: An improved CryoSat-2 sea ice freeboard retrieval algorithm through the use of waveform fitting, The Cryosphere, 8, 1217–1237,doi:10.5194/tc-8-1217-2014, 2014.

Kwok, R., Cunningham, G. F., Wensnahan, M., Rigor, I., Zwally, H. J., and Yi, D.: Thinning and volume loss of the Arctic Ocean sea ice cover: 2003–2008, J. Geophys. Res., 114, C07005, doi:10.1029/2009JC005312, 2009.

Laxon, S.: Seasonal and interannual variations in Antarctic sea ice extent as mapped by radar altimetry, Geophys. Res. Lett., 17, 1553–1556, doi:10.1029/GL017i010p01553, 1990.

Laxon, S.: Sea-ice altimeter processing scheme at the EODC, Int. J. Remote Sens., 15, 915–924, 1994.

Laxon, S., Peacock, N., and Smith, D.: High interannual variability of sea ice thickness in the Arctic region, Nature, 425, 947–950, doi:10.1038/nature02050, 2003.

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.: CryoSat-2 estimates of Arctic sea ice thickness and volume, Geophys. Res. Lett., 40, 732–737, doi:10.1002/grl.50193, 2013.

Lundhaug, M.: ERS SAR studies of sea ice signatures in the Pechora Sea and Kara Sea region, Can. J. Remote Sens., 28, 114–127, 2002.

Ricker, R., Hendricks, S., Helm, V., Skourup, H., and Davidson, M.: Sensitivity of CryoSat-2 Arctic sea-ice freeboard and thickness on radar-waveform interpretation, The Cryosphere, 8, 1607–1622, doi:10.5194/tc-8-1607-2014, 2014.

Scagliola, M.: CryoSat footprints, Aresys technical note, SAR-CRY2-TEN-6331,Aresys/ESA,Italy, 2013

Scagliola, M.: Main evolutions and expected quality improvements in SAR/SARin BaselineC Level1b products, Aresys technical document,C2-TN-ARS-GS-5154, Aresys/ESA, Italy, 2014.

SIGRID3 Manual: SIGRID-3: a vector archive format for sea ice georeferenced information and data, JCOMM Technical Report No. 23, Tech. rep., World Meteorological Organization, Geneva, Switzerland, 2014.

Similä, M., Mäkynen, M., Cheng, B., and Rinne, E.: Multisensor data and thermodynamic sea-ice model based sea-ice thickness chart with application to the Kara Sea, Arctic Russia, Ann. Glaciol., 54, 241–252, doi:10.3189/2013AoG62A163, 2013.

Soussi, B. and Femenias, P.: ENVISAT RA-2/MWR Level 2 User Manual, ESA User Manual, ESA, ESRIN, Italy, 2006.

Vidard, A., Balmaseda, M., and Anderson, D.: Assimilation of altimeter data in the ECMWF Ocean Analysis System 3, Mon. Weather Rev., 137, 1393–1408, doi:10.1175/2008MWR2668.1, 2009.

Wingham, D., Rapley, C., and Griffiths, H.: New techniques in satellite altimeter tracking systems, in: ESA Proceedings of the 1986 International Geoscience and Remote Sensing Symposium (IGARSS '86) on Remote Sensing: Today's Solutions for Tomorrow's Information Needs, 8–11 September 1986, Zurich, Switzerland, 3, 1339–1344, 1986

Wingham, D., Francis, C., Baker, S., Bouzinac, C., Brockley, D., Cullen, R., de Chateau-Thierry, P., Laxon, S., Mallow, U., Mavrocordatos, C., Phalippou, L., Ratier, G., Rey, L., Rostan, F., Viau, P., and Wallis, D.: CryoSat: a mission to determine the fluctuations in Earth's land and marine ice fields, Adv. Space Res., 37, 841–871, doi:10.1016/j.asr.2005.07.027, 2006.

Zygmuntowska, M.: Arctic sea ice altimetry – advances and current uncertainties, PhD thesis, Dept. Geophysics, Univ. Bergen, Bergen, Norway, 2014.

Zygmuntowska, M., Khvorostovsky, K., Helm, V., and Sandven, S.: Waveform classification of airborne synthetic aperture radar altimeter over Arctic sea ice, The Cryosphere, 7, 1315–1324, doi:10.5194/tc-7-1315-2013, 2013.

Table 1. Classification Matrix of CS-2 Based Classification (rows) and AARI Ice Chart (columns), November 2013. Last column shows the best and the worst hit rates of the three 5-day periods.

	FY < 70 cm	MY	Open ~~Water~~ Ocean	Worst–Best
FY < 70	51 %	46 %	3 %	31–64 %
MY	8 %	92 %	0 %	88–97 %
Open ~~Water~~ Ocean	5 %	4 %	92 %	87–95 %

Table 2. Classification Matrix of CS-2 Based Classification (rows) and AARI Ice Chart (columns), March 2014. Last column shows the best and the worst hit rates of the three 5-day periods.

	FY $<$ 70 cm	FY $>$ 70 cm	MY	Open ~~Water~~ Ocean	Worst–Best
FY $<$ 70	20 %	60 %	7 %	13 %	15–26 %
FY $>$ 70	1 %	84 %	14 %	0 %	75–91 %
MY	0 %	14 %	86 %	0 %	77–92 %
Open ~~Water~~ Ocean	2 %	1 %	3 %	93 %	90–98 %

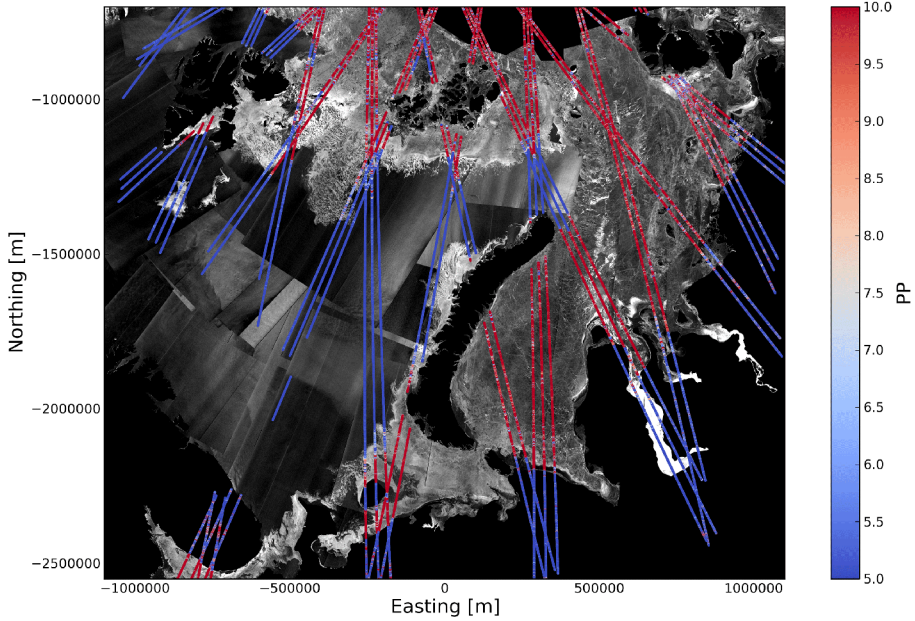

Figure 1. CS-2 Pulse Peakiness (1–5 March 2014) plotted on a RS-2 SAR composite (5 March 2014).

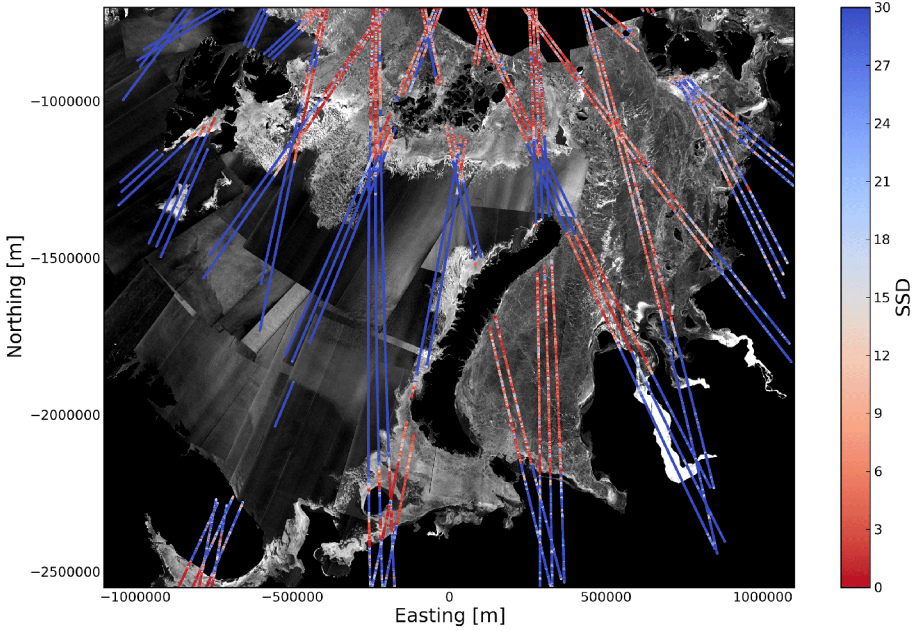

Figure 2. CS-2 Stack Standard Deviation (1–5 March 2014) plotted on a RS-2 SAR composite (5 March 2014).

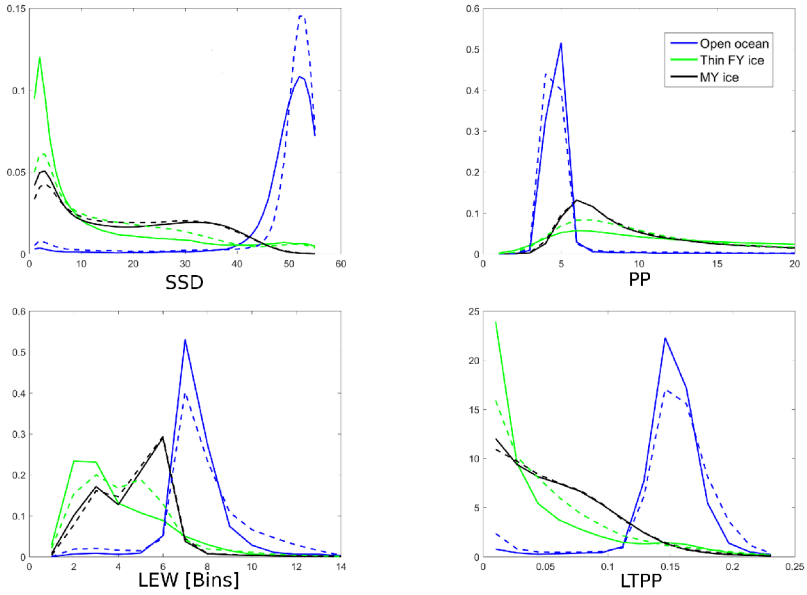

Figure 3. Normed distributions of waveform parameters: SSD (top left), PP (top right), LeW (low left) and ~~TPP~~ LTPP (low right) for the period 1–15 November (solid lines) and 16–30 November (dashed lines) 2013.

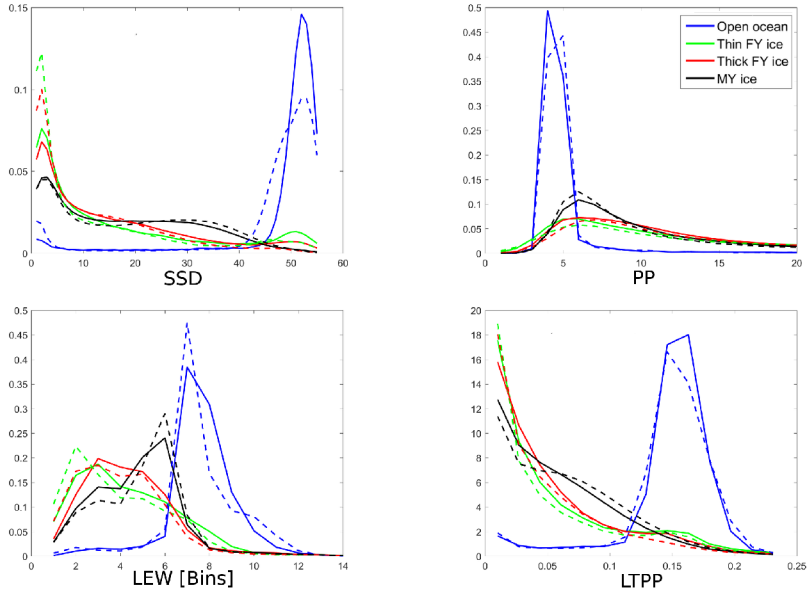

Figure 4. Normed distributions of waveform parameters: SSD (top left), PP (top right), LeW (low left) and ~~TPP~~ LTPP (low right) for the period 1–15 March (solid lines) and 16–30 March (dashed lines) 2014.

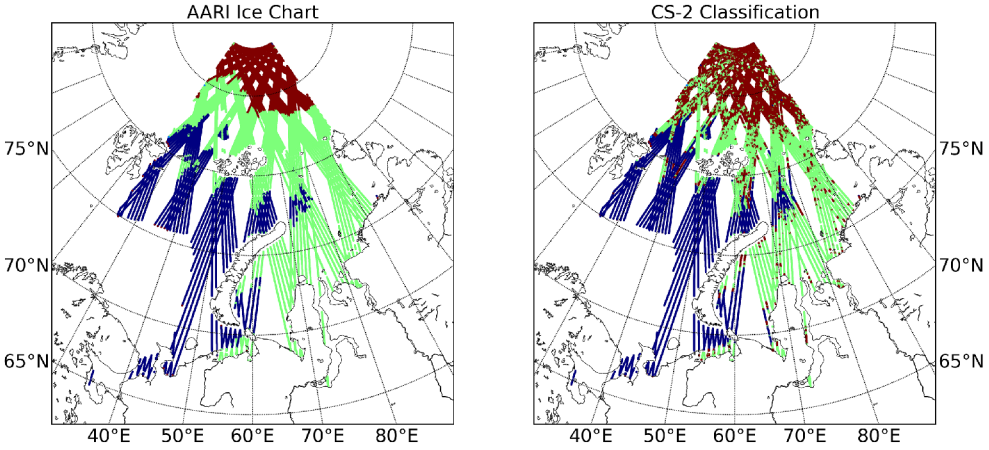

Figure 5. Automatic classification test for ~~March 2014.~~ November 2013. AARI Ice Chart SA sampled at CS-2 footprints (left) and the classification result from CS-2 measurements (right). 15–30 November 2013. Blue = ~~Open water~~open ocean, green = FY < 70 cm and Red = MY.

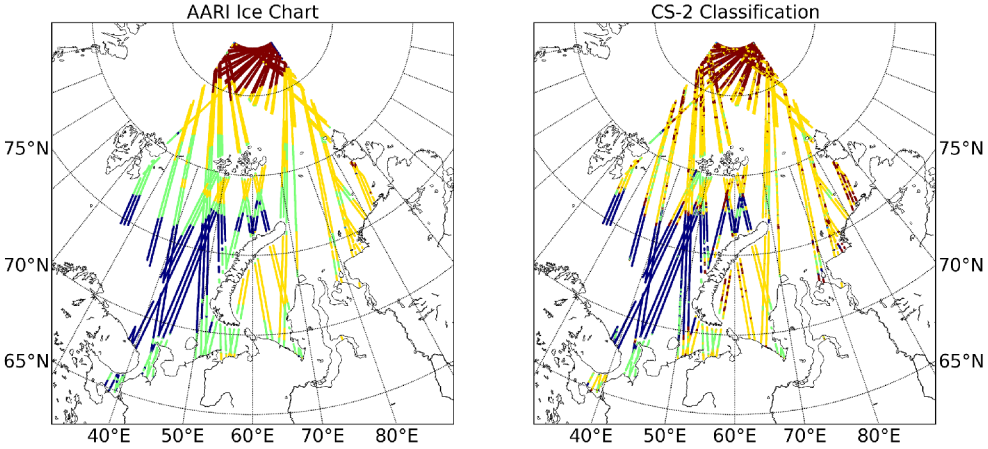

Figure 6. Automatic classification test for March 2014. AARI Ice Chart SA sampled at CS-2 footprints (left) and the classification result from CS-2 measurements (right). 15–30 March 2014. Blue = ~~Open water~~open ocean, green = FY < 70 cm, yellow = FY > 70 cm and red = MY.