

We thank the reviewers for their comments and constructive criticism, which have helped us to improve the manuscript. Below follow our answers and comments, all written in *italics*.

Comments to Referee #1

The Cryosphere Discuss., 7, C1269–C1270, 2013
www.the-cryosphere-discuss.net/7/C1269/2013/
© Author(s) 2013. This work is distributed under
the Creative Commons Attribute 3.0 License.

Interactive comment on “Comparison of automatic segmentation of full polarimetric SAR sea ice images with manually drawn ice charts” by M.-A. N. Moen et al.

Anonymous Referee #1

Received and published: 1 August 2013

Overall, this is a good article that attempts to answer the very difficult problem of automated image classification for sea ice. The authors present their findings in a clear and concise manner with a sound methodology. Listed below are my specific comments and technical corrections:

1. Although the authors address the manual charts (completed by 2 different ice analysts), I am surprised that there is such a difference between both products. A couple of things come to mind: a) Is there a difference in the analysts' experience (ie. How long they have been doing ice analysis)? b) Could the disagreement between the manual charts be attributed to the fact that the analysis was done on RGB composite and a Pauli decomposition products? The authors mention that the analysts are experienced in using R-2 ScanSAR (probably single channel HH or dual-channel HH/HV) and optical data in their daily chart creation. Therefore, they are likely more comfortable analyzing grey-scale images and have less experience interpreting these colour SAR products. I wonder if the manual charts would have been more consistent if the SAR data had been presented as greyscale images.

Answer 1:

The two analysts had 16 and 2 years of experience in ice chart analysis. During the manual ice chart analysis they were able to turn on/off separate channels in order to access grey scale images. For more information see Answer 17.

The following corrections have been made:

- *Added sentence*
p.2612 – l.19: “One of the analysts had sixteen years of experience drawing ice charts and the other one just two.”
- *Added sentence*
p.2602 – l.23: “The analysts were able to access the separate channels as grey-scale images by switching off channels.”

2. I assume the decision to use 6 features is described in Doulgeris and Eltoft,2010. It may be useful to include a couple of sentences describing this decision.

Answer 2:

Following corrections have been made:

- *Added sentences*
p.2604 – l.5: “Six empirical real-valued features were extracted from the covariance matrix using the Extended Polarimetric Feature Space (EPFS) method (Doulgeris and Eltoft, 2010 and Doulgeris, 2013). Five features are basic polarimetric parameters known to characterize the polarimetric signature of the illuminated area. This standard feature space has been extended to include a non-Gaussianity feature. All features have shown good potential in segmentation and most of them have a reasonable physical interpretation. More information about the features can be found in Drinkwater, 1992 and Doulgeris, 2013.”
 - *Included new reference, Doulgeris, A. P.: A simple and extendable segmentation method for multi-polarisation SAR images, in Proc. POLinSAR 2013 – 6th Int. Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry, vol. ESA SP-713, pp. 8, Frascati, Italy, in Press, 2013.*
3. Can the author clarify the comment on Pg 5-Ln 450: "From a SAR imaging point of view, it is not possible to separate all these classes by visual inspection of RGB images from polarimetric channel combinations." For first-year ice types, MANICE identifies 4 types: (first-year (Code 6), thin fy (Code 7), med fy (Code 1.), thick fy (4.). National ice services regularly distinguish these first-year types from SAR data on their ice chart products. I recognize that the analysts in this study use different names for their firstyear classes. Perhaps it would be useful to use the MANICE definitions rather than the types listed in Fig.4.

Answer 3:

The following corrections have been made:

- *We have adopted the classes and codes defined by WMO (Sea Ice Nomenclature, No.259 Supplement 4, 1989)*
- *The reference to the WMO Sea Ice Nomenclature has been included in the reference list.*
All classes used in the manual ice charts are now according to this document. The WMO codes do not distinguish between Nilas with and without frost flowers. This difference is clearly detectable in SAR images, thus the additional class “Nilas with frost flowers” has been included.
- *Changed sentences*
p.2608 – l.3ff: “By taking the labels into consideration, we noticed that the yellow (class 7) and all the green labels (classes 8–11) describe various stages of first-year ice. The segments labeled first-year ice at different stages are the areas with the most differences. Separating especially the first-year ice classes based on visual interpretation of intensity SAR images only seems to be a very subjective part of manual ice charting. Merging all the first-year ice classes would make the ice charts more alike.”
- *Added sentence:*

p.2608 – l.7: “Ice services that use SAR for their SoD ice chart products distinguish the different first-year ice types, but not without the help of additional data such as coastal and shipboard ice observations and knowledge of the ice development history. Such additional data are not always available, which is why the code “First Year” (code 6) is sometimes used. This code does not specify a sub-range of thicknesses within the definition of first-year ice, as opposed to codes 7, 8, 9, 1• and 4•”

4. Technical Correction: Pg 2-Ln 64. R-2 ScanSAR Wide imagery is not 50m resolution. The beam mode has 50m pixel spacing. Resolution for SCW is 160-72.1m (Range) x 100m (Azimuth).

Answer 4:

The official RS-2 website confirms that referee #1 is correct.

- *The following correction has been made:*

p.2597 – l.18: “Radarsat-2 ScanSAR Wide scenes have a coverage of 500 km x 500 km with 160 – 72.1 m (ground range) x 100 m (azimuth) resolution.”

5. Spelling Correction: Pg 5-Ln 396. Change to "fourteen"
The misspelling has been corrected
6. Spelling Correction: Pg 8-Ln 692. Change to "largest"
The misspelling has been corrected

Comments to Referee #2

The Cryosphere Discuss., 7, C1400–C1405, 2013
www.the-cryosphere-discuss.net/7/C1400/2013/
© Author(s) 2013. This work is distributed under
the Creative Commons Attribute 3.0 License.

Interactive comment on “Comparison of automatic segmentation of full polarimetric SAR sea ice images with manually drawn ice charts” by M.-A. N. Moen et al.

Anonymous Referee #2

Received and published: 12 August 2013

General comments

This is a well-written paper that presents an interesting compilation of data, combining SAR polarimetry, ship observations, airborne ice thickness measurements and manual classifications by ice chart analysts. The three aims of the paper and the broader justification for this work are sound and highly relevant with increasing ship traffic in the Arctic and the need for more accurate ice charts. Also, extracting quantitative information about ice type and ice morphology from SAR polarimetry is an important goal in of itself. Hence, the type of comprehensive analysis advanced in this paper is both timely and relevant.

Unfortunately, the manuscript falls short of its declared goals in a number of respects that are outlined in detail below. These points need to be addressed in revising the paper.

In its present form the study is interesting but runs the danger of doing a disservice to both ice charting and polarimetric SAR analysis by failing to establish a coherent, transparent framework for the comparison between the manual and automated classification approaches.

(1) There is lack of clarity with respect to the comparison between the automated classification of the SAR data and the manual classification carried out by ice chart analysts. In the title of the paper and the introduction (and other parts of the text) it is implied that the study compares ice charts by two analysts with polarimetric data segmented objectively. As pointed out by the authors on p. 2602, ice charts are generated using a multitude of data sources; here specific reference is made to SAR data and “available optical data”.

However, in the description of methods it is unclear whether the analysts relied on remote sensing data other than the Radarsat SAR data. At the same time, they utilized ship-based observations and photographs, information which may only rarely be available for standard classification in the context of production of operational ice charts.

Answer 5:

The phrase referred to on p. 2602 (“The Norwegian Ice Service’s operational ice charts are manually drawn based on dual-pol ScanSAR Wide data and available optical data”) concerns operational ice charts only. The intention was to stress the difference between what the ice analysts usually do (interpretation of dual-pol and optical data), and what they were told to do for this particular study. We have moved this sentence, see bullet point number 4.

In this paper we have used the phrase “Ice chart” as a general expression for any ice chart, not just those drawn operationally.

The following corrections have been made:

- *Changed /added sentence to emphasize that no operational charts are produced in our study:
p. 2599 – l.1: “None of the charts were intended for operational use.*
- *Added sentence to highlight that the ice charts developed for this study are non-operational.
p.2602 – l.1: “In this section we present the methods used for the preparation of the manual ice charts and the automatic segmentation of the SAR data, all were non-operational ice charts.”*
- *Corrected sentence:
p.2602 – l.18: “The authors would like to stress that the ice analysts have less experience in using quad-pol SAR scenes for ice type classification.”*
- *Moved the first paragraph, p.2602 – l.7-10 to l.18 to have less focus on the operational charts.*

Answer 6:

Section 2 lists the available data used in this study, no other remote sensing data except Radarsat-2 were accessible. Section 3.1 describes how the manual ice charts were produced and all ancillary data utilized.

The following corrections have been made:

- *Added p.2602 – l.24*
“No other data was available for the following analysis”.

Since the paper outlines in the introduction that the research is motivated by an improvement of operational ice charts, more detail needs to be provided on the manual segmentation process and data sources used. For example, if “available optical data” were used as in standard charting operations (such as AVHRR, MODIS or DMSP OLS) then the type of data and their spatial resolution and extent need to be clearly stated. What about thermal IR data, which will be highly effective in distinguishing ice types at the surface temperatures observed in the region. If such data was used, how did cloudiness impact their use across the scene?

See answer 6.

Also, the level of experience of the two analysts needs to be clearly stated as well.

See answer to referee #1 part 1).

If on the other hand, the analysts did not generate a standard ice chart but simply used their expertise in manually segmenting the SAR scene by relying on surface based observations and photographs, then this needs to be made clear.

If such an approach was taken, then the title and key sections of the paper would need to be revised to avoid any confusion of a manual classification with the generation of an actual ice chart.

At present section 4 of the paper and the discussion are difficult to evaluate since critical information is missing. Some wording in Section 4 (e.g., on p. 2609) or Section 5 (p. 2612) implies that in fact the analysts prepared standard ice charts.

However, if that is the case, then much of the discussion and key conclusions of the paper are of limited value with respect to aims #1 and #2 outlined on p. 2599.

With the corrections described in Answer 5 we hope to have clarified this point.

Specifically, the paper does not make it clear what the polarimetric classification criteria are compared against in the way of datasets entering into the manual classification. Moreover, the match between the different charts cannot be compared rigorously without a more detailed evaluation of the optimal number of classes required to describe the observed ice types.

Answer 7:

We are not fully sure what referee #2 means by this.

The polarimetric classification is performed using the standard Mixture of Gaussian (MoG), using features explained in the manuscript. These features are different from the information the analysts viewed, in particular the features may enhance aspects not easily seen by eye.

One of our conclusions states that “The number of classes is a critical input parameter which constrains the algorithm”. The current study is part of an ongoing project and we have plans for further investigation of the “optimal number of classes”.

(2) The number of predefined classes is a critical element of the comparison between manual and automated classification. Since the ice chart analysts appeared to have had

access to a much broader range of data, they can be expected to arrive at a larger number of classes than the fully automated algorithm.

Answer 8:

A broader range of data does not necessarily lead to recognition of a larger number of classes. We have shown that the automated algorithm can distinguish at least as many classes as the analysts. The main question is the class distinction in the data. As stated in Answer 7, the features may enhance aspects which are not visible by eye.

Answer 9 gives our reasoning in the choice of the number of predescribed classes for the automated algorithm. We would like to stress that this choice applies for this study which only looks at snow-covered first-year ice in spring. Ice conditions with other ice types present (e.g. multiyear ice, ponded ice, heavily deformed ice) will require a new assessment of the optimal number of classes; equally, the ice analysts will likely arrive at a different number of classes than in the present study.

However, to limit the number of classes in the automated analysis to 5 based on a subjective approach is problematic. Rather, in this type of unsupervised classification ideally objective criteria need to be employed to determine the optimal number of classes. Such criteria can be based on measures of covariance, as determined from a principal components analysis. Alternatively, the number of classes could have been constrained based on expert judgement by the ice analysts to match their evaluation.

Lacking such detail, it is difficult to assess the significance of the differences between the three independent segmentation approaches discussed in the paper.

Answer 9:

We agree that the number of classes is a critical and complicated issue. Several optimization criteria exist, but no general criteria applicable on all kinds of data. These criteria are not very well suited for SAR images, especially for detailed sea ice images like ours. The methods we have been using are state-of-the art methods shown useful for sea ice SAR images.

The automatic algorithm finds the number of classes it is told to find; we tested the algorithm by using different number of classes as input. These tests indicated that increasing the number of classes as input to the automatic algorithm did not make the segmentations seem more similar. See p.2615 – l.11ff.

We are aware that the “optimal number of classes” is likely to vary with which expert you ask and the applied method. The number of classes would probably also change with the intended application of the ice chart.

In our scene the ice analysts found 6 and 4 classes (p.2629 – fig.4), while the ice experts (those attending the cruise and also were doing the field work) only recognized 3 based on optical images from the helicopter flight, and 4 classes by also including the Pauli image (table 2). According to the shipboard ice log 6 classes were observed at the acquisition time of the RS-2 image. Based on that information we found 5 to be a reasonable number of classes. However, our results revealed that one of the classes should be split (see p.2615 – l.7ff). This will be dealt with in further on-going studies.

(3) The discussion of the range of different polarimetric parameters for different ice classes and open water on p. 2613 and 2614 assumes that there is in fact a comparatively narrow, unique set of parameters that describes each of these different ice types. However, neither

for open water and nor for thicker first-year ice is this likely to be the case. As is well established (and discussed in some of the papers referenced by authors), backscatter signatures over open water depend on wind speed, fetch and wind direction relative to the viewing geometry. For first-year ice, it is the type, orientation and distribution of deformation features (ridges, ice rubble, rafting etc.) that will determine SAR backscatter signatures. These aspects may not be reflected in differences between the statistics of different ice thickness distributions (and likely also won't show up in visible/thermal IR satellite imagery). This raises the question as to whether the findings from this study can be interpreted in more general terms without a more detailed analysis of the ice morphology or the wave spectrum developed over open water.

Answer 10:

We agree with the statements reviewer #2 presents. In the referred paragraph, "Polarimetric parameters" on p. 2613 – 2614 we summarize our results and compare them with other relevant studies. Well aware that the consequence of altering conditions (e.g. weather and viewing geometry) during image acquisition is dissimilar results/measurements, no conclusions are drawn. However, some features may be invariant to some changing conditions. This will be considered in our next study, a comparison of segmentations from three consecutive days (including the quad-pol scene considered in this paper).

The following corrections have been made:

- *p.2614 – l.21: "Well aware that altering conditions (e.g. weather and viewing geometry) during image acquisition will change the backscattered signal, thus making image interpretation more complicated, we do not draw any conclusions relating mean value and standard deviation of each class and feature to an actual ice type. Some features may be invariant to changes in the viewing geometry (e.g. incidence angle), this needs to be further investigated in future studies."*

(4) The introduction of the paper nicely made a case for improving automated approaches that can support the generation of ice charts for a range of applications. The conclusions need to be more specific in regards to steps that need to be taken for such improvements to take effect. For example, in several places the paper emphasizes how the ice analysts were not familiar with polarimetric SAR products. Does that mean that a simple training course in the interpretation of polarimetric SAR data can vastly improve the quality of operational ice charts? If so, what steps need to be taken towards that goal?

Answer 11:

At present, quad-pol SAR images are essentially used for site studies, not for operational ice charting. This is mainly due to the poor spatial coverage per image (25 km x 25 km). Our intention was not to improve the "manual" part of manual ice charting (by further educating ice analysts). We wanted, firstly, to investigate the benefits of polarimetry for sea ice charting, and thus comprehend the limitations of dual-pol. Secondly, to explore how polarimetric information can be utilized in operational sea ice services. I.e. how can polarimetry increase the efficiency of the manual classification process? Our study has shown that it can be used for both segmentation and physical interpretation of the segments (labeling), but this need to be further investigated (see answer 12).

Along the same lines, to what an extent can one extrapolate from the current study to conditions that are much more representative of typical shipping operations in the Arctic in summer and fall? At that time, warm, wet ice surfaces may mask many of the differences between ice types observed in late winter and early spring.

Answer 12:

As previously discussed (see answer 10), changing weather conditions during image acquisition will change the backscattered signal, thus making the image interpretation more complicated. Polarimetric features are not considered invariant through seasonal changes. One possible solution can be to use knowledge of the ice developing history and classified ice charts from previous consecutive days as priori knowledge. This information can be used to develop a so-called trained classifier.

The following corrections have been made:

- *p.2515 –l.15: “Polarimetric features are not expected to be invariant through seasonal changes. Thus the classes and the interpretation of the features acquired in one season are not directly transferable to scenes from other seasons. A possible solution is to use a priori information, such as knowledge of the ice history and ice charts from previous days to develop a so-called trained classifier.”*

Specific comments

P. 2596 - l. 13-14: This sentence is vague and sounds more like an advertisement than a conclusion from the paper. Please be more specific (and appended directly to previous sentence in abstract without paragraph break).

The following corrections have been made:

- *Changed sentence*
p. 2596 – l.13-14: “...used in the segmentation were interpreted in terms of physical sea ice properties. Utilizing polarimetric information in sea ice charting will increase the efficiency and exactness of the maps. The number of classes used in the segmentation has shown to be of significant importance. Thus, studies of automatic and robust estimation of the number of ice classes in SAR sea ice scenes will be highly relevant for future work.”

- l. 16: “Arctic sea ice cover” is more specific than “Arctic ice cover” unless you are also referring to lake and riverice.

The following corrections have been made:

- *Changed sentence*
p. 2596 – l.16: “The Arctic sea ice cover has changed significantly during the last decades”

- l. 19: Increasing human activities such as shipping are driven as much by other factors such as global economics or geopolitics as by changes in the ice cover, please reword.

The following corrections have been made:

- *Sentence is changed to:*
“This development is a contributing factor to the observed increase in shipping and exploration activity in ice infested Arctic areas.”

p. 2597 - l. 25: "processes"
Misspelling is corrected

p. 2598 - l 8ff: I assume you are referring exclusively to SAR data here since there are a number of studies that have compared ice charts based on visible range and IR range imagery and passive microwave products to other types of validation data. Please be specific. If you are referring to broader validation efforts then these other studies (authored, e.g., by researchers at the CIS or the NIC in the US) need to be referenced.
The following corrections have been made:

- *Changed sentence:*
"There is not much work published on the validation of manual ice classification charts or on pixel-to-pixel comparisons between manual charts and automatic segmentations utilizing SAR data exclusively."

p. 2599 - l. 8: "relative kurtosis" in the context as used here is cryptic, please explain briefly – i.e., does this refer to the fourth moment of a distribution of a polarimetric variable derived over a specific region or is this jargon for something else?

Answer 13:

The "relative kurtosis" is related to the normal interpretation of kurtosis, which is always positive. There are several definitions of kurtosis. The "excess kurtosis" subtracts the "absolute kurtosis" of Gaussian data so the "excess kurtosis" of the standard normal distribution is zero. This definition gives uniform distributions a negative (excess) kurtosis and sharp peaked distributions a positive (excess) kurtosis. We have divided (instead of subtracting) the "absolute kurtosis" by the absolute kurtosis of Gaussian data; hence we call it the "relative kurtosis", and "RK" equals one for normal distributions. "Flat" distributions (relative to the Gaussian distributions) will have "RK" less than one, "peaked" distributions (relative to the Gaussian distribution) will have "RK" above one.

The following corrections have been made:

- *Added sentences:*
p.2599 – l.8: "see section 3.2"
- *p. 2604 – section "Relative Kurtosis:*
"The relative kurtosis (RK) is a statistical measure of the shape of the distribution. It is the absolute kurtosis measure divided by the absolute kurtosis of a Gaussian distribution, hence the name "relative" kurtosis. Normalizing by division falls naturally out of the product model, commonly used for radar texture modeling (Oliver and Quegan 2004, Doulgeris 2013). Distributions with high kurtosis are characterized by having a sharp peak close to the mean, a rapid decline and heavy tails relative to Gaussian data."

p 2600 - l. 5: The Pauli colour coding scheme needs to be explained in the figure caption as well, or better yet, the three RGB channels need to be shown in the figure legend along with the corresponding channel combination.

The following corrections have been made:

- *Figure caption has been changed.*

p.2626 – Fig. 2a: “Radarsat-2 scene, 12 April 2011. Geocoded polarimetry image shown as Pauli colours (the intensity channel combinations $|HH-VV|$, $2|HV|$ and $|HH+VV|$ are assigned to the RGB channels, respectively).”

p. 2601 - l. 14: “acquiring”
Misspelling is corrected

p. 2603 - l. 16: What is the rationale for a 21x21 scattering vector matrix? Is this driven by the desired effective resolution or chosen for other (computational?) reasons?

Would one expect to see similar classification results for a much larger or smaller number of scattering vectors?

Answer 14:

The average processing described by eq.(1), p.2603 is accomplished for several reasons. It will reduce speckle noise, thus increase the radiometric distinction between different regions/classes, which is desired. On the other hand the pixel resolution will be degraded causing a blurred image with less detail. One consequence is a smoother/cleaner segmentation which is easier to interpret. Another advantage is of course, as referee # 2 mentions, the computational benefit. Averaging over 21 x 21 scattering vectors turned out to be a reasonable compromise for this particular image. The following corrections have been made:

- *Added sentence:*

p.2603 – l.17: “...over L number of scattering vectors. This process needs to balance several conflicting requirements, e.g. speckle noise reduction, increase of radiometric distinction, degraded pixel resolution. In this study, $L = 21 \times 21 = 441$, which is a reasonable compromise for this dataset.”

p. 2604 - equation 3 and l. 10ff: The text is a bit confusing here (e.g., please clarify “heavy tail” in terms of the pdf at several standard deviations away from the mean). Please also explain how specifically RK relates to the true kurtosis of a Gaussian distribution and what you mean by “non-Gaussianity”. Since RK is always non-zero and positive I’m not entirely sure I see how this relates to the true kurtosis.

Answer 15:

When we talk about “relative kurtosis”, we always mean relative (ratioed) to the normal distribution. “Heavy tail” here refers to heavier than the tails of the Gaussian distribution. Since all pdfs integrate to one, a narrow distribution with a sharp peak must have “heavier tails” than a broader, more rounded, Gaussian distribution with the same mean value. The relative kurtosis is a measure of “non-Gaussianity”. It measures how similar/dissimilar a given distribution is compared to the Gaussian distribution. All phrases in question are commonly used statistical terms.

Changes in the manuscript regarding “relative kurtosis” are described in Answer 13.

p. 2606 l. 19: “fourteen”
Corrected

p. 2607 - l. 2: Please use standard nomenclature when referring to what appears to be a rank-order filter.

Answer 16:

We have used standard nomenclature in the manuscript. A “rank-order filter” is not the same as a “majority filter”.

p. 2612 -l. 7 ff: The reference to “the ice analysts have used too many classes” makes little sense and points to a key flaw in the analysis. If the ice analysts drew on the variety of data sources hinted at further up in the paper, then the segmentation into 7 or 6 classes is probably very well justified.

However, comparing two sets of classifications based on different premises and datasets requires further work to attribute discrepancies between different classifications.

Answer 17:

We agree that the wording has not been clear enough. The ice analysts arrived at 4 and 6 classes, so the number of classes they arrived at is probably not too large. However, we think the reason for the discrepancy between the two manual ice charts is mainly due to the use of different first year ice classes. Separating the sub-classes of first-year ice by merely visual interpretation is the most subjective part of manual ice charting (see answer 3 and 9).

The following corrections have been made:

- *Added/changed sentence:
p.2612 – l.7: “We believe that the main reason for the dissimilarity between the manual segmentations is the various labels of first-year ice. I.e. subjectivity appears to be an important factor in labeling the first-year ice in particular. The manual charts would probably be more similar if the number of first-year ice labels were reduced.”*
- *Added sentence:
p.2614. – l.14: “Even the two manual charts are inconsistent to some degree. However, the difference is mainly due to the analysts using the first-year ice classes differently. This supports the idea of sea ice charting being subjective.”*

p. 2618 -l. 32: correct spelling of Nghiem

Corrected