Response to the comments of reviewer #2 to

**Assessment of error in satellite derived lead fraction in Arctic**
by N. Ivanova, P. Rampal, and S. Bouillon

In the following we provide detailed answers (marked by "**A**") to your comments ("**R**") and describe what changes are done in the revised manuscript ("**Changes**").

**R: General comments**
In general, the paper is well written and structured, and easy to read and understand. The data processing and analysis methods are scientifically sound and discussed in needed detail.

**A:** Thank you for this positive feedback on our work, we appreciate it very much.

**R:** My main critic and concern in the paper is the method used for SAR based estimation of LF:
1) It is very simple, not up to current state-of-art in SAR based sea ice classification. The algorithm is basically a backscattering coefficient threshold method to separate leads with calm open ocean or thin ice (low sigma0) from sea ice (high sigma0). Leads with rough ocean or deformed thin ice are not identified, as the authors point out. The threshold method results a binary mask of leads vs. sea ice at 100 m pixel size which can be aggregated over a larger area to LF estimates. The threshold algorithm uses as input median filtered (5x5 pixel window) sigma0 values, which reduces the effect of radar speckle, but I think that still the radar speckle has some effect on the results if a sigma0 value is close to threshold causing randomness to which class, lead or sea ice, a pixel is assigned.
I suggest that the authors estimate the equivalent number of looks in their SAR imagery rectified to 100 m pixel size, yielding an estimate for radiometric resolution of the sigma0 (~10*log10(1+1/sqrt(ENL))), and study possible effects in the lead-sea ice classification.
The authors claim that no method has so far have been presented in literature addressing automatic LF retrievals from SAR. I would say that this may not be true as recent years papers have published on SAR based sea ice concentration (SIC) retrieval which may also be applied here, see e.g.:
Anders Berg and Leif E. B. Eriksson, "SAR Algorithm for Sea Ice Concentration—Evaluation for the Baltic Sea", IEEE Geoscience and Remote Sensing Letters, Vol. 9 (2012), 5, p. 938 - 942.
Karvonen, J.: Baltic Sea ice concentration estimation based on C-band HH-Polarized SAR data, IEEE J. Sel. Top. Appl., 5, 1874-1884, doi:10.1109/JSTARS.2012.2209199, 2012.
Karvonen, J.: Baltic Sea ice concentration estimation based on C-band Dual-Polarized SAR data, IEEE T. Geosci. Remote, 52, 5558-5566, doi:10.1109/TGRS.2013.2290331, 2014. 2
Steven Leigh, Zhijie Wang, David Clausi. "Automated Ice-Water Classification Using Dual Polarization SAR Satellite Imagery," IEEE Transactions on Geoscience and Remote Sensing, vol 52, no 9, 2014.
Liu, Huiying; Guo, Huadong; Zhang, Lu, "SVM-Based Sea Ice Classification Using Textural Features and Concentration From RADARSAT-2 Dual-Pol ScanSAR Data", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,

8(4), pp 1601-1613, 2015.
I suggest that the authors conduct a thorough review of SAR based sea ice classification methods relevant to their study, and see if they can develop or apply a better method for the LF estimation. If you decide to continue use your current method then a firm detailed justification is needed.

**A:** The SAR-based LF retrieval method used in the paper is indeed very simplistic, and we understand the concern raised by the reviewer. We agree that we might have been not enough clear in explaining why we think the method can be used as it is for the goals set in this work. Therefore here and in the revised manuscript we will provide a more clear and detailed justification, as suggested by the reviewer. The two main points raised by the reviewer are: a) more sophisticated and up to current state-of-the-art SAR sea ice classification methods, available from literature, are not used; b) the effect of the speckle noise on the classification.
a) As the reviewer has pointed out there have been a few SAR-based sea ice classification methods published. What we wanted to say in our manuscript was that none of them has been applied to perform automatic lead fraction retrieval. We believe that every specific application of SAR-imagery processing requires individual approach. This stems from the fact that SAR provides indirect information about the surface properties we are interested in, which requires an individual approach for conversion of this indirect information (backscatter) to each physical parameter. Therefore one cannot assume that the mentioned methods will work for the case of identifying leads. For example, Korosov et al. (2015) have demonstrated that these thin features cannot be distinguished using an SVM-based method applied to SAR subsets (windows), while such a technique was good enough for ice/water separation in general. Also this study showed that even applying this method to segments, which significantly improved the resolving capacity of the method, was not good enough, and it was found that the SVM needed to be trained specifically targeting leads. The same would apply to the method suggested by Berg and Eriksson (2012) where a neural network would need to be trained to fit the purpose of lead fraction retrieval and to be suitable for the Arctic sea ice, which is different in properties from that of the Baltic Sea. The method of Karvonen (2012) is also developed for the Baltic Sea, and its aim is to retrieve sea ice concentration, so it is not designed to separate open water from thin ice, which is important when working with leads in Arctic winter. The methods by Karvonen (2014), Leigh et al. (2014) and Liu et al. (2015) are all designed for dual-polarization SAR images, which we do not have access to. Sentinel-1 with its free access to the SAR images was not considered because of lack of overlap in time between Sentinel-1 (launched in 2014) and AMSR-E (stopped operating in 2011). We agree that it is possible to implement these methods, adapt them for our purpose and test their performance, but a task like this belongs to method development for SAR-based lead fraction retrieval, which is out of scope of this study. It would deserve a separate study and publication, and would definitely have been of great value. The focus of our paper is passive microwave retrieval of lead fraction. Such a dataset is highly valuable and therefore we set a goal to put uncertainty bars on it to the best of our knowledge and to look for possibilities to improve it. We consider our SAR lead fraction dataset to be good enough for this purpose for the reasons explained in the following answers.

b) Speckle noise is sufficiently reduced by the median filter applied in the suggested method, and the features are enhanced by it, as demonstrated in the Fig. 3a, b of the manuscript. The panel b of the figure shows that the features are outlined quite precisely. From our visual inspection of the resulting retrievals (including comparison to MODIS) we saw that the noise did not affect the outcome largely. It is however still possible that a few pixels on the border of each feature are indeed misclassified – something we cannot capture with our eyes – and an uncertainty is introduced. However, if we take into account the fact that the AMSR-E-based method is expected to resolve relatively wide leads (> 3 km, Röhrs et al., 2012) and the retrieval has a grid size of 6.25 km × 6.25 km, while the grid size of our SAR-based retrieval is 100 m × 100 m, we assume this uncertainty is small in comparison to the difference between the AMSR-E and SAR retrievals especially for the bin of lead fractions near 100% for AMSR-E, which represent the main issue with the dataset. Another way speckle noise could affect the retrieval is providing single-pixel (or very small cluster of pixels) false leads, but such cases are excluded from the study by excluding AMSR-E LF < 1%. Such cases were not observed in the considered SAR imagery either: if SAR had such a small feature identified as lead, AMSR-E did not have a value there. And if AMSR-E had had a value, this would have indicated that it was not a result of speckle noise in the SAR image, but a true feature.

**Changes:** In the revised version of the manuscript we will provide the clarifications given above (shortened) and include an overview of the publications on SAR sea ice classification methods in the Introduction.

**R:** 2) The AMSR-E LF product accuracy is studied using an another LF product based on satellite data, this case SAR, and further, the SAR based method itself is not fully validated. The authors constructed a manually quality controlled sub-set of SAR based LF data for the AMSR-E LF accuracy studies. I suggest to study if you could add visual spectrum MODIS and MERIS data from March-April to further validate your SAR based LF estimation. At least larger leads should be visible in the optical imagery. During nighttime also 1 km MODIS ice surface temperature data could also useful (e.g. as by Willmes and Heinemann 2015). My work with these data have shown that manual quality control of cloud masking is unfortunately needed. The authors could also check availability of fine resolution Landsat etc. data over their study area.
In general, I am concerned about validating one remote sensing product with another product, which itself is fully mature and validated. Your study is more about intercomparison of two RS products and finding their differences, and you may not get solid information about the AMSR-E LF product accuracy against the true (in-situ) lead fraction. You claim that AMSR-E LF product is overestimating LF, but are you sure that your SAR based LF method is not underestimating, and thus, leading to this conclusion?

**A:** The reviewer questions the reliability of the reference SAR lead fraction dataset and suggests using MODIS, MERIS or Landsat products to validate the reference dataset further. As a consequence, the reviewer also points out that with our data we should not be able to quantify the errors in the AMSR-E dataset. Therefore, we would like to provide a clear step-by-step description of our approach, which will explain why we think that our reference dataset is acceptable for the purpose and the error estimations are useful. We might have been unclear in our original manuscript in this regard.

a) We conclude that the AMSR-E method overestimates lead fraction first from the visual analysis of the overlays of image pairs – SAR image and AMSR-E lead fraction retrieval – an example is shown in the Fig. 5 of the manuscript (the zoom-in) and explained in the text. We refer particularly to the last bin of the AMSR-E LF distribution, which contains large amount of observations close to 100% (Fig. 1 of the manuscript), while in SAR images the area covered by leads in such grid cells was obviously smaller in almost all the cases. The cases where one lead width would take the full AMSR-E grid cell or even more (lead width larger than one grid cell) were extremely rare in our selection.

b) In order to quantify this overestimation we retrieve LF from SAR. The SAR-based approach may indeed underestimate lead fraction, but we performed visual inspection of every SAR image, which indicated that we detected most of the leads. In addition, quality control is performed to exclude the ambiguous cases (see below). The residual underestimation seemed much smaller in comparison to the overestimation by AMSR-E clearly seen from the visual analysis of image pair overlays (AMSR-E LF and SAR images).

c) We make an assumption that the grouping of large amount of lead fraction values around 100% in the AMSR-E dataset (Fig. 1 of the manuscript) can be fixed by dividing the incorrect values by a certain factor. But some values are assigned 100% correctly, and since we do not know which ones, we approach the problem from the other side. We imitate the problem by multiplying the lead fraction retrieved from SAR (in all bins) instead, and we retrieve a distribution similar to that of the original AMSR-E dataset, which confirms our assumption. We find the best match for such factor varying from 2.5 to 3.7, but to allow for some uncertainty, we say it is between 2 and 4. So, already here we account for possible underestimation/overestimation by SAR.

d) The last clarification that needs to be provided here is the quality control of our SAR-based reference dataset.

We will first mention that we have performed a visual comparison of the SAR dataset to a MODIS image, and the SAR method captured the majority of features correctly. However the time difference between the images was 12 hours, and thus some of the features have changed their width. This will be an issue in general, if one would like to compare lead fraction from SAR and MODIS – it is extremely hard to find sufficient amount of image pairs which will be close in time, cover the same spatial area, have leads present in the footprint, and be cloud-free (MODIS) – all at the same time. In our visual comparison of this one pair of images there were indeed some smaller features that were missed by the SAR method. We chose however to restrain ourselves from performing of a validation of our SAR dataset versus MODIS (MERIS or Landsat), because quantification of lead fraction from such measurements requires another threshold technique, which will need validation too.

Therefore we did following to check the quality of the reference dataset:

- Making sure that our main dataset provides the same distribution as high quality dataset where every subset was visually analyzed and only certain cases selected (please see the description of the MQC SAR LF dataset).
- Visual evaluation of every triple of images: original SAR image, SAR lead fraction, AMSR-E lead fraction. Ambiguous cases were discarded (please find the details in the manuscript).
- We were only considering the AMSR-E grid cells (6.25 km by 6.25 km size) that had LF value > 1% and where the SAR LF for this grid cell returned a value of LF > 1% too. This means that we have excluded all the leads the SAR has eventually

missed. This is what we mean by that our SAR dataset is tailored to the purpose. It is not dependent on the AMSR-E dataset, but we only validate AMSR-E grid cells, which contain a non-zero lead fraction value.

**Changes:** We will make the description of our approach clearer in the revised version of the manuscript according to what is provided above.

**R:** 3) SAR based LF estimation is difficult, especially using single channel SAR imagery as in your case (ENVISAR C-band HH-pol WSM). I propose that you check if finer resolution ENVISAT APP images are available over your study are in Nov 2008 – Apr 2009. If that is the case then use APP based LF estimates to validate those from coarser resolution WSM imagery. In addition, are there any RADARSAT-2 dual-pol ScanSAR images available (through MyOcean or metno's Ice Service data archive)? Addition of cross-pol data could enhance lead detection.
How about ICESat data for SAR lead detection validation? ICESat data were acquired in Nov 2008 – Apr 2009?

**A:** For the time and area of interest we had neither APP or dual-pol images available. The ICESat data in the winter 2008–2009 were of low quality (K. Khvorostovsky, radar and laser expert at NERSC, personal communication) and were not used by other experts in the field either (e.g., Kwok et al:
http://rkwok.jpl.nasa.gov/icesat/download.html; Yi and Zwally:
http://nsidc.org/data/nsidc-0393).

**R:** In summary, I would like to see you study further accuracy of your SAR LF estimation method, as it is new one and presented for the first time in your paper.

**A:** Above we have provided justifications on the choices we made with regard to the accuracy of the SAR-based method. Here we would only like to add that this work was not aiming at development of a method for lead fraction retrieval from SAR, but rather focused on validating of non-zero measurements available in the AMSR-E dataset.

**R:** In the paper you mention/discuss about thin ice, but you don't define exact thickness range or WMO ice classes for the thin ice. 3

**A:** Since we were evaluating the existing lead fraction dataset from AMSR-E, we follow the definition of thin ice given in the original paper of Röhrs et al. (2012). There thin ice is defined according to World Meteorological Organization (WMO, 1989) as including new ice, nilas, and pancake ice.

**Changes:** the definition of thin ice will be included in the revised manuscript.

**R: Specific comments**
Page and line numbers refer here to the printer-friendly version of the article.

**Introduction**

**R:** Page 6317, line 1: "Model simulations showed that even 1% change in sea ice concentration due to the increase in areal lead fraction can lead to a 3.5 K difference

in the surface temperature"
Does this refer to sea ice surface temperature or surface air temperature?
**A:** Near-surface atmospheric temperature is meant here.
**Changes:** Adjusted accordingly.

**R:** p. 6317, l. 17: "Accurate observations of lead fraction are thus of high importance for model evaluation and for being assimilated into models as initial conditions, or during a simulation.
How about the exact lead locations, especially larger ones, are they important?
**A:** From a purely forecast perspective, we investigated that question and found that having the right location of leads at time zero is definitely important if one wants to get e.g. the right drift pattern in the following few days. However, the statistics of drift and deformation can be captured by only inserting statistical information about the leads like their amount and their sizes, almost independently of their location. This was only tested so far on simulations ran over the central Arctic, not for regions covering only the marginal seas like e.g. the Barents Sea.

**R:** l. 6318, l. 16: "However, this approach is limited in time coverage because AMSR2 started to deliver the data only in 2012 (http://suzaku.eorc.jaxa.jp), and quantitative validation work is still needed."
Quantitative validation work is needed - is this your opinion or found in a journal paper?

**A:** The paper of Beitsch et al. (2014) where this method is presented is targeting sea ice concentration and is not adapted to lead detection as for example is done (by deliberate enhancement of thin elongated features) in Röhrs et al. (2012) using similar instrument (AMSR-E). On the other hand, we understand the concern of the reviewer – the AMSR2 resolution used in the Beitsch et al. (2014) is finer than that of AMSR-E and this may mean that the enhancement is not necessary. However, our opinion is that quantitative validation is needed because Beitsch et al. (2014) only do visual assessment using three consecutive MODIS images (true color and ice surface temperature), which gives qualitative assessment, but no number reflecting the difference between the sea ice concentration at 3 km resolution and the MODIS image is obtained. A quantitative assessment is only provided for sea ice concentrations retrieved by different algorithms and at different resolutions from passive microwave data.

**Changes:** This point will be clarified in the Introduction.

**R:** p. 6318, l. 28: "The selected classifier was able to detect 68% of leads correctly, and only 3% of ice measurements were falsely identified as leads."
Why quite high fraction of leads were missed with CryoSat-2 data? This could be of interest to a reader.

**A:** Since our work is not addressing methods of lead fraction retrievals from CryoSat-2, we would prefer not to go into much detail on this aspect. We believe it is a combined result of several reasons, and interested readers can refer to the original paper of Wernecke and Kaleschke (2015), which we cite in the manuscript.

**R:** p. 6319, l. 26: "Based on analysis of the errors we introduce a correction factor for

the existing dataset and suggest an improvement of the AMSR-E based method itself."
The AMSR-E LF data cannot be corrected, but you propose a tie point correction to the LF algorithm itself? Please correct me, if I am wrong.
**A:** This is correct.
**Changes:** Adjusted accordingly.

**2.1 The AMSR-E LF dataset**

**R:** p. 6320, l. 12: "LF is expressed as the percentage of a grid cell covered by leads, which are represented by either open water or thin ice."
What is the thickness range of thin ice here?
Please explain shortly in this sub-Section the method for the AMSR-E based LF estimation.
The AMSR-E LF is a daily gridded product? This is not explicitly mentioned.
**A:** On the thickness please see our answer in the end of the general comment. Yes, it is a daily gridded product.
**Changes:** We will add 1-2 sentences describing the method (some details are already provided in the Introduction), and add that it is a daily gridded product.

**2.1 The SAR images**

**R:** p. 6321, l. 19: "The SAR images originally provided with spatial resolution of 75m.75m,"
This is pixel size of the WSM images, not the resolution. Provide resolution and noise floor, available from ESA docs. Investigate many equivalent of number looks are in your rectified SAR images, see my comments above.
**A:** Geometric resolution is 150 m ×150 m, pixel spacing 75 m × 75 m. On the noise, please see our reply to the general comment 1), with the noise floor the same reasoning applies.
**Changes:** The resolution will be specified.

**R:** p. 6321, l. 22: "Calibrated surface backscattering coefficient (ASAR Product Handbook, 2007) normalized over ice was used for this study (we will refer to this value as backscatter)."
Unclear what this normalization means, incidence angle variation compensation?
**A:** yes, it is empirical incidence angle variation compensation.
**Changes:** Will be specified in the manuscript.

**3 SAR-based threshold technique**

**R:** p. 6322, l. 22: "The threshold was first tried with nd=1 and nd = 2 (dashed red lines), but it was found that an intermediate value nd = 1.5 (solid red line) worked better and therefore was chosen."
How this was determined? Visually analyzing results with different nd? 4
**A:** Yes, visual analysis comparing the lead fraction retrievals with different threshold values.
**Changes:** Will be specified in the manuscript.

**R:** p. 6323, l. 3: "This method is developed strictly for the purpose of the AMSR-E

LF dataset validation and therefore does not represent an independent LF retrieval method from SAR."
I don't understand this, to my opinion your SAR LF should be independent of the AMSR-E LF dataset if you want to validate the AMSR-E LF.
What is the effect of SAR noise floor in your sigma0 threshold based LF method? Measured sigma0 is sum of target true sigma0 and SAR noise equivalent sigma0. Can the threshold in (1) be lower than the noise equivalent sigma0? If so then leads are not detected at all.
**A:** This is a problem of our formulation; we will adjust it in the revised version. What is meant here is that we only validate AMSR-E grid cells where a non-zero lead fraction is available. We also discard all the SAR images where the classification did not work because we are not developing a SAR-based method and therefore we do not need to make it work for all possible scenarios. We take each non-zero AMSR-E grid cell and "look" what is in there in a SAR image. The procedure is performed automatically on many images.

We know that the leads are detected because we can see it visually. The leads that are missed by SAR do not play any role because such areas are excluded (they do not participate in the comparison with AMSR-E). Mixed cases where only part of a lead is identified by SAR and the rest is not (which will make this AMSR-E grid cell eligible for the comparison), has almost not been encountered. Please note that the AMSR-E method is identifying only relatively wide leads (>3 km), for such size of the leads SAR was either identifying the whole lead or not identifying anything, but no partial identification has occurred.

**Changes:** The issue will be clarified in the manuscript.

**4.1.1 MQC SAR LF**

**R:** p. 6324, l. 13: "Defining a threshold locally not only eliminates significance of these effects, but it takes advantage also of less variety of surfaces in general."
How threshold definition was done, manually?
**A:** yes.
**Changes:** Will be clarified in the manuscript.

**R:** p. 6324, l. 20: "The classification in each subset was then inspected visually, comparing the three collocated maps: backscatter, MQC SAR LF and AMSR-E LF, in order to make sure it was successful."
Now your MQC SAR LF dataset is not independent of AMSR-E LF? If so then how this effects your SAR LF vs. AMSR-E LF comparison? I think your SAR LF should be totally independent of AMSR-E LF.
**A:** The MQC SAR LF dataset is not dependent on AMSR-E LF. This comparison of the triplets of the images was done with main focus on the checking of the MQC SAR LF against the original backscatter, while we also looked at the AMSR-E LF solely in order to exclude triplets where for example large areas of AMSR-E LF subset were missing. This last part was only for purposes of the dataset cleaning.
**Changes:** Will be clarified in the manuscript.

**4.1.2 SAR LF**

**R:** p. 6325, l. 5: "The majority of subsets contained leads represented by signatures darker than surrounding background, while if those with brighter signature were present in large amount such images were discarded."

This is a serious drawback of your SAR based LF method, leads with high sigma0 are not detected. You have circumvented this by discarding SAR images with leads having high sigma0. Another, and better, way would be to further develop your algorithm.

**A:** Yes, if our purpose were to develop a SAR-based lead fraction retrieval method, this would have to be done in a better way.

**5 Discussion**

**R:** p. 6238, l. 20: "A method to retrieve LF from SAR backscattering coefficient is introduced. This simple threshold technique is only suitable for the purposes of this study, and is thus not universal. However, its potential is shown, and the limitations are identified, which allows further developments of the method."

It is good that you admit many limitations of your SAR LF method, but I am not very happy about circumventing this by stating that it is still suitable for your study. You could target more universal SAR LF algorithm to increase value of your paper. See my General comments.

**A:** Please see our answers above.

**R:** p. 6329, l .22: "When the distribution is bimodal (one mode for leads and one for thicker ice), a value between the peaks can be used as threshold, as suggested by Lindsay and Rothrock (1995) for distributions of temperature or brightness. However, such cases were so rare in the selected SAR images that this approach was discarded."

Yes, also to my experience cases when a sigma0 has a clear bimodal distribution are rare, as there are leads with low or high sigma0, smooth FYI with low sigma0 etc., i.e. the overall sea ice sigma0 distribution can have many local peaks.

**R:** p. 6331, l. 2: "For example, it would not be able to capture leads narrower than 3 km due to its resolution,...

This limitation should also mentioned in Introductory and 2.1 The AMSR-E LF dataset Sections.

**A:** It is mentioned in the Introduction (p. 6317 l. 28).

**References:**

*(References listed in the manuscript are not duplicated here)*

Korosov, A., Zakhvatkina, N., and Muckenhuber, S. Ice/water classification of Sentinel-1 images. EGU 2015, 17 April 2015.

WMO: The World Meteorological Organization Sea Ice Nomenclature (WMO No. 259, TP-145, Supplement No. 5), 1989.