Author Response

We want to thank the Referees for their reviews which helped us a lot to revise and (we think) improve the manuscript. The main concern was the resolution of MODIS and its limits for the identification of narrow leads. We analyzed the spread of the SSH which is smaller for a threshold on the maximum power of the waveform than for all classifiers compared with. We hope that this result helps to dispel some concerns and feel that it is a great complement to the original manuscript.

Referee 1

P2168L4: suggest "comparison" instead of "a combination".
Corrected

P2168L16: suggest "strongly reduces" instead of "is strongly reducing". This statement isn't necessarily true for momentum – see Martin et al. (2014), "Seasonality and long-term trend of Arctic Ocean surface stress in a model", JGR Oceans, and what they found about the 'optimal sea ice concentration'. I would say something like: "Sea ice strong modifies air-ocean interaction etc …"
You are right sea ice does not reduce momentum transports in all cases, but only if a compact sea ice cover is present. This is typically the case outside of the marginal ice zone. Taking into account that the other mentioned transports (heat and mass) between ocean and atmosphere are always reduced by sea ice we consider the statement (which includes the word 'most') as true. As 'reduce' (correctly) stresses the importance of leads more than 'modify' we prefer to keep using 'reduce' but in the suggested way.

P2168L20-25: What is the resulting heat balance of leads? There is loss of latent heat as new ice is formed, but more absorption of solar radiation – what is the net effect?
The latent heat flux per lead area is largest in winter due to the higher temperature differences. The absorption of solar radiation is largest (as mentioned in the opposite direction) in summer, as it is directly connected to the incoming solar radiation. The latter has a larger impact on the annual mean net flux and is expected to play a mayor role in the Arctic amplification (e.g. Yoshimori et al., 2013). However it would be misleading to compare their amplitudes as they appear at different times and the lead area is likely to change from summer to winter.

P2169L16-17: Do you have a citation for this statement?
A citation (Andreas et al., 1979) has been added

P2169L18: Suggest "dampens" rather than "is dampening".

Corrected

P2169L28: Suggest "utilise" rather than "are utilising".
Corrected

P2170L1: Are the authors aware of the paper by Renner et al. (2014), "Evidence of Arctic sea ice thinning from direct observations", GRL? It is not true that studies using airborne data are limited to just individual seasons.
You are right, the results of individual campaigns can clearly be combined to investigate temporal changes (like also done in Renner et al., 2013). The mentioned citation has been added and the statement clarified. It is still a regional study.

P2170L16-18: Laxon et al. (2013) showed excellent agreement between CS2 and in situ data. I understand what you mean by this sentence but to say that altimeter estimates are not "satisfying" is unfair. I think it is more appropriate to say that knowledge of the snow loading and the radar interaction with the snow layer currently limits the accuracy of altimeter derived sea ice thickness estimates.
Corrected

P2170L23: Reverse this sentence i.e. "SSH is crucial for altimeter based ice thickness retrievals"
Done

P2170L26-28: What do you mean by thin ice? Ice <~10cm thick will have a fairly negligible effect on SSH retrieval as freeboard will be ~1cm and speckle noise >10cm.
A SSH bias due to thin ice is in contrast to speckle noise not reduced by averaging (multi-looking), as it is always positive. The effect on ice thickness estimates is comparable to the mean ice thickness of lead measurements. If this is e.g. ~5cm we get a bias of a few percent of the derived ice thickness (for typical thicknesses) which we consider as small but noteworthy.

P2171L14-16: Do you have a citation for this statement? Not everyone in the altimeter sea ice community will know anything about MODIS interpretation so a citation would be useful.
The word 'sufficient' has been removed as this is now discussed separately. A citation is given.

P2171L18: "this" should be "these"
Corrected

P2171L12-P2172L2: For the ground truth dataset, I worry about the effect of very small leads. It is known that even very small amounts (even ~1% of the footprint area) of open water can cause specular reflections that will dominate the echo [Drinkwater, 1991], but would these

small leads show up in MODIS data? Maybe using MODIS data means that the analysis is valid for leads over a certain size, but there might be some caveats for smaller leads?

The following parts have been added:

'Note that this method is limited by the resolution of MODIS. CryoSat-2 measurements which look like originating from ice in MODIS scenes can actually contain small amounts of leads. See section 4.2 for a discussion of this circumstance.'

'It has been shown that leads which cover only a small fraction of a radar altimeter footprint can dominate the signal, due to the high amplitude of specular returns (Drinkwater, 1991). Therefore CS-2 detects leads which are simply not visible for MODIS despite its higher resolution. The fraction of this leads in the ice class of the ground truth cannot be quantified by our approach. Those narrow leads either cover the nadir point or not, while leads covering the whole footprint ('True Leads') do for sure. Therefore one could expect True Lead measurements to ensure a higher quality (see section 4.3) for the derivation of the SSH.

This expectation is supported by the smaller spread of the SSH estimate based on the MAX_1 compared to the PP_1 with nearly the same amount of lead detections (True + False Leads, Table 1). This advantage is on the other hand certifying that narrow, unnoticed leads in the ice class do not reverse the ROC analysis.'

In addition the focus on the False Lead Rate has been reduced, e.g. by replacing its mention it in the abstract and conclusion.

P2172L3-P2173L7: You should provide some references/citations for this section.
Done

P2173L4-7: This isn't a very clear explanation. This is just summing the power in each beam and fitting a Gaussian. Maybe explain how the summed beam power relates to/varies with specular or rough surfaces.
This part has been revised and the expected behavior of the SSD and SK has been added in section 2.3.

P2173L16: I don't think that Laxon et al. (2013) give an explicit definition of pulse peakiness (PP). The PP of Peacock and Laxon (2004) is multiplied by 31.5 so I don't know where the factor of 100 comes from. I think you should be careful when you refer to "the Laxon et al. (2013) classification" throughout the paper as the classification is not explicitly stated in Laxon et al. (2013).
You are right, the PP definition used by Laxon et al. (2013) is not stated clearly at all. A

threshold of 18 in combination with a factor of 31.5 does not detect any leads. Kurtz et al. (2014) state (P1227): 'In this study, we used the pulse peakiness and stack standard deviation thresholds used by Laxon et al. (2013)[...]' and they use a PP threshold of 0.18 instead of 18. We trust this conversion as Nathan Kurtz is co-author of Laxon et al. (2013).

We have to mention, that the PP definition of Kurtz et al. (2014) is different to our and those by Peacock and Laxon (2004), Laxon (1994) and Armitage and Davidson (2014) (which is the reference within Kurtz et al. (2014)) as it is written as sum(1/P_i) instead of 1/(sum(P_i)). This has also been noted in the review of John Padon (http://www.the-cryosphere-discuss.net/8/C74/2014/tcd-8-C74-2014-supplement.pdf) but has not been corrected (or mentioned) in the final version. We take this as a typing error, like Robert Ricker has confirmed that it is in Ricker et al. (2014) (Robert Ricker, personal communication, January 2015).

We have introduced abbreviations for the classifiers used by Laxon et al. (2013) and Ricker et al. (2014) and avoid the mentioned term as well as terms like 'defined by Laxon et al. (2013)'.

P2174L16-19: The trailing edge of CS2 waveforms follows a 1/sqrt(t) shape, not exponential [Wingham et al, 2006]
This statement is, for our understanding, not completely accurate. Inserting the last case of A2 in A1 yields to 1/sqrt(t)*exp(-a*t), which is also illustrated in Fig. 15a in Wingham et al. (2006).

, so why the choice of an exponential fit?
The short answer is that 1/sqrt(t) has no parameter describing the scale at which the trailing edge of the waveform drops down. In addition we maintain consistency with other studies (e.g. Zygmuntowska et al., 2013).

P217420-P2175L3: Again, this isn't a particularly clear explanation of the beam behaviour parameters. Refer to the fact that a Gaussian function is used to approximate the summed power in each beam
Done

P2175L4-10: Describe what the kurtosis represents intuitively i.e. the 'peakedness' of a distribution. Not all readers will be aware of this.
Done

P2175L12-P2176L11: This seems like a good methodology, although I am unfamiliar with it myself. I also imagine that a lot of readers in the sea ice altimetry community are unfamiliar with this methodology as well. Please provide some references in this section for interested

readers.
Done

I would also encourage you to provide a slightly longer and more detailed explanation here. For example, it is not clear how you derive the thresholds in the training subset. Perhaps it would be instructive to walk through the methodology in more detail for a given example (e.g. PP with w=0.05) and include one or two figures.

The thresholds are derived by minimizing the cost function on the training set. This is done for each classifier separately. In the two parameter case both thresholds are derived at the same time (the combination of both with the smallest value of the cost function is considered as optimal). As the cost function is dependent of w, different thresholds are found for different w. We revised the explanation and added a flow chart.

P2177L2-5: What do these waveforms look like? If they are specular what else (apart from leads) could have caused them? This links back to my comments about the resolution of the MODIS data, it's ability to detect very small leads and the sensitivity of altimeters to very small leads. Section 3.2: Would it be possible to quantify the classification performance in terms of the retrieved elevation? For altimetry, and in particular for estimating sea ice thickness, we are actually interested in the SSH determination. Perhaps for the best performing classification parameters and thresholds the authors could compare the elevation transects? Another way SSH precision is usually quantified is to calculate the along track variance of the SSH at 20Hz, 40H, 100Hz. This would be an excellent addition to the paper. Whilst the authors have shown that lead detection is of interest in itself (for lead distribution and lead width studies), this paper would be of a much broader interest to the altimetric community if the classification parameters were assessed based on their ability to measure SSH and not just detect leads in the first place.
An analysis of the SSH has been performed and the issue of very small leads discussed. Each measurement detected as lead by the classifier used by Laxon et al. (2013) is to some amount specular, as it has a PP>0.18
The question whether this is a good threshold is part of this study and we can not answer it on a theoretical basis.

P2179L3-14: Do the authors have an explanation for the apparent discrepancy between CS2 and AMSR-E? It seems that CS2 is detecting many more leads in the Marginal Ice Zone – perhaps CS2 is more sensitive to smaller leads that are not detected by AMSR-E?
We agree, see P2184L6-14

Section 3.3: Surely CS2 will consistently over estimate the lead width? The CS2 tracks will not always be orthogonal to the leads and when CS2 crosses leads at oblique angles the lead width as you have defined it will always be over estimated.

We agree, that's why we distinguish (in contrast to earlier studies) between apparent and actual lead width. Section 4.5 now states that the apparent lead width is typically larger than the actual.

P2181L8-14: The 'false lead rate' again leads me to question whether the grounds truth you have presented are simply missing small leads, that could still dominate the CS2 return. If the waveforms are being identified as specular then there must be a very flat surface within the pulse-limited footprint (i.e. not sea ice), so what other explanations can the authors think of for these echoes? Also, it seems that these classifiers are performing very similarly in terms of the lead detection rates, but I would stress again that what really matters to CS2 users is the SSH determination. How do the MAX and PP classifiers perform when it comes to SSH determination? I would strongly encourage the authors to include this in the study.
We agree and included such analysis.

P2182L20-P2183L3: This paragraph is very unclear and I do not really understand what you are saying. Suggest it is re-written more clearly.
Done

P2183L5-6: This is not a 'possibility' as you say here - it has been shown to be true.
Corrected.

Section 4.2: Whilst I think that you may be correct here, this section is all just vague speculation. Again, I encourage the authors to include some comparison of the elevation transects using the optimised classification threshold, and/or the along track SSH variance. This is an easy step to take and you would then be able to say for sure whether the MAX classifier is better than PP at getting rid of off-nadir leads. This is what is of interest to the altimetry community.
Yes, this section is very speculative (all but the last sentence is speculated by Armitage and Davidson (2014)). Due to the importance of off-nadir leads and the physically reasonable argumentation, we feel that it is still worth mentioning. Especially as it is now supported by the SSH estimate.

P2184L14-19: The ice edge from CS2 seems to extend right into the central Norwegian Sea, which isn't very realistic – so there seems to be some lead detection where there is no ice present?
In Fig. 5a all grid cells with at least 2000 measurements are displays. This is the case in the central Norwegian Sea but none of this measurements is detected as lead (lead fraction=0, colored in black)

P2185L5-7: I am not convinced by this explanation – it is known that leads can dominate the return if they cover a few percent of the footprint.

The following, alternative explanation has been added:
'The MAX_1 is optimized mainly on leads wider than a single measurement which could also cause the relative small number of apparent lead width of 300 m.'

Section 4.3 & 4.4: Again, surely the CS2 lead width and spatial distribution will be affected by the fact that CS2 will not necessarily cross leads orthogonally?
This is the case for the lead width but for our understanding not for the lead fraction.

Lead fraction:
Imagine for example parallel leads. The width of the ice between the leads appears in the same way larger than the width of the leads. We see no reason why this should be different in more realistic scenarios. Each measurement can be regarded separately. It is classified as lead or ice, depending on the surface properties within its footprint. This is not influenced by the track orientation.

Is it possible to account for this effect or correct for it in some way?
Lead width:
As mentioned in section 4.5, a transformation of the lead width is not possible without further assumptions. This has not been done so far (all previous studies also consider the apparent lead width distribution) and it is not necessary for some applications like the turbulent heat transport. It is clearly stated, that this is not the actual lead width i.e. the short dimension of a lead but the width how it appears to CS2.

P2187L16-19: You haven't shown this to be true – it might be the case that using the different parameters and thresholds you have presented makes no difference to the SSH determination, or it might be that the MAX parameter is worse than PP. I don't think that this is necessarily the case, but you could easily show if it is true or not by including a comparison of SSH using the optimised lead classifications parameters.
Corrected

Figure 2: It would be instructive to plot the across track extent of the CS2 pulse limited footprint if possible – this would give an indication of where off-nadir leads are dominating the return.
This would in deed be instructive on a smaller scale (like in Fig.5 in Armitage and Davidson (2014)). On this scale (the Y-Axis has a length of about 200km) some markers 750m to the left and right of the track would be obstructive to compare the visible leads with the CS2 detections. As the main purpose of this figure is to show the differences in lead detections (and (a) to visualize the method) and not the influence of off-nadir leads, we feel that the clearness of the illustration prevails. The resolution of MODIS is not sufficient to correctly visualize the presence of most off-nadir leads anyways.

Andreas, E. L., Paulson, C. A., William, R. M., Lindsay, R. W., & Businger, J. A. (1979). The turbulent heat flux from Arctic leads. Boundary-Layer Meteorology, 17(1), 57-91.

Yoshimori, M., Watanabe, M., Abe-Ouchi, A., Shiogama, H., & Ogura, T. (2014). Relative contribution of feedback processes to Arctic amplification of temperature change in MIROC GCM. Climate dynamics, 42(5-6), 1613-1630.

Referee 2

Does it mean, that leads of smaller size (<250 m) are not considered?
It basically does. We expect to see some indications for a lead in MODIS even if it is not covering the whole pixel. In this case the CS-2 measurements have been excluded from the analysis. However we do agree that leads which are too narrow to be seen by MODIS can be correctly detected by CS-2 which reduces the value of the ROC analysis. Those small leads either cover the nadir point or not, while leads covering the whole footprint ('True Leads') do for sure. As we are not able to distinguish between nadir and off-nadir leads in the first case, it is comprehensible to privilege the latter. The validity of the optimization analysis is supported by the reduced SSH variance of MAX_1 leads to those from other classifiers (e.g. PP_1). We interpret this reduced variance as a sign for less ice measurements being detected as lead and/or less off-nadir leads.

In any case, uncertainties of the "ground truth" and limits of the statistical analysis should be discussed in more detail.
The following parts have been added:

'Note that this method is limited by the resolution of MODIS. CryoSat-2 measurements which look like originating from ice in MODIS scenes can actually contain small amounts of leads. See section 4.2 for a discussion of this circumstance.'

'It has been shown that leads which cover only a small fraction of a radar altimeter footprint can dominate the signal, due to the high amplitude of specular returns (Drinkwater, 1991). Therefore CS-2 detects leads which are simply not visible for MODIS despite its higher resolution. The fraction of this leads in the ice class of the ground truth cannot be quantified by our approach. Those narrow leads either cover the nadir point or not, while leads covering the whole footprint ('True Leads') do for sure. Therefore one could expect True Lead measurements to ensure a higher quality (see section 4.3) for the derivation of the SSH.

This expectation is supported by the smaller spread of the SSH estimate based on the MAX_1 compared to the PP_1 with nearly the same amount of lead detections (True + False Leads, Table 1). This advantage is on the other hand certifying that narrow, unnoticed leads in the ice class do not reverse the ROC analysis.'

'Deficiencies of the ground truth which might be caused by ice drift and opening/closing of leads between the data acquisition, collocation and unnoticed narrow leads increase the error rates which might therefore be overestimated.'

The authors further state, that all measurements with a mixture of both classes within the footprint are excluded. But a lead could be still detected if the CryoSat-2 footprint contains both classes and therefore be a valuable tie point, given that it is not off-nadir.
Those measurements are excluded from the ground truth as they would complicate the threshold optimization. The ground truth is not used in context of the lead width and fraction estimates as well as for the SSH, where all measurements are included with the previously optimized classifiers.
It is now clearly stated, that there might be some narrow leads in the ice which cannot be seen by MODIS and can therefore not be excluded from the optimization.

Detailed comments:

P2168 L15: "affects" instead of "modulates" seems to fit better here.
corrected

P2169 L15: may be replace "Not only" by "Apart from the lead area, also the".
corrected

P2170 L10-L18: I think here it needs a better clarification of the term "freeboard" with respect to laser and radar altimetry.
We now introduce the terms 'snow freeboard' and 'ice freeboard'

P2170 L12-13: may be replace "into an ice thickness in hydrostatic equilibrium" by "into ice thickness by assuming hydrostatic equilibrium".
Done

P2170 L14: from Ku band radar altimetry.
corrected

P2170 L21: It should be stated that the interferometric mode is not performed across the

entire Arctic, but only in the "Wingham Box" (which is now abrogated) and the coastal areas.
Done

P2171 L4: Some short introduction/description of MODIS is required here from my point of view.
Has been added

What is the minimum size of a lead that can be detected with MODIS images?
We are not aware of any study which answers this question for the visual (250m resolution) bands. We do speculate that there is some sign of a lead even if the pixel is covered completely. However we can not prove this and therefore have to assume that the minimum size is the resolution of 250m.

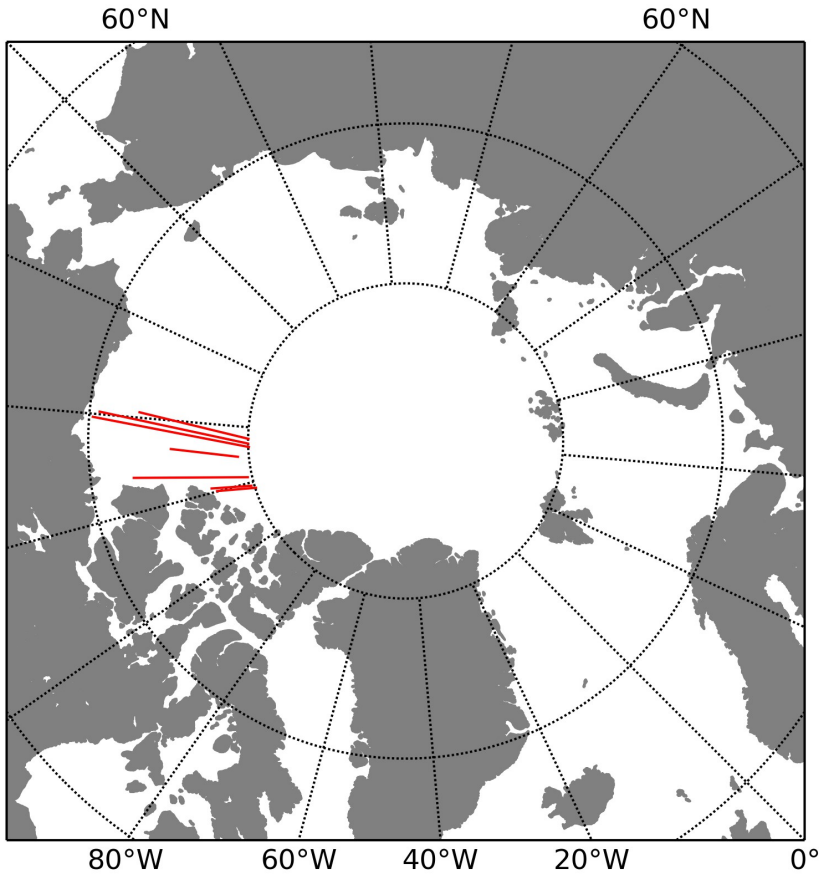
P2171 L14: Can the authors provide a reference and short description of the MODIS data used in this study?
Done

P2171 L20-21: Does this mean that if when MODIS shows a lead but also any fraction of ice within a CS-2 footprint, the measurement is excluded?
Yes

Wouldn't this only allow leads of a remarkable size?
Yes, this allows only leads which are large enough to cover the CS-2 footprint

This needs some clarification. which size is assumed for the CS-2 footprint?
The assumed footprint size (300 in and 1500m across track) has been added.

P2171 L24: Can the authors provide information about the location and record time of the used MODIS granules? An additional map (may be incorporated in one of the other figures) would be beneficial.
The granules are quite large because of which the location of them is not very instructive. Therefore we plotted the positions of the CS-2 tracks within those granules. We would prefer not to incorporate a map in Figure 3 (which is the only one really related to the ground truth) as it would reduce its size and we feel that it is already on its maximum regarding the amount of provided information. Including this map as a new figure would on the other hand be of very limited interest for most readers. A description of the measurement positions is added to section 2.1



The start times (UTC) of the 5 min acquisition intervals is provided in the following:

Terra	March	05 2013; 22:20
Terra	March	12 2013; 22:25
Terra	March	03 2013; 22:30
Terra	March	16 2013; 22:00
Terra	February	28 2013; 22:00
Aqua	May	02 2012; 12:25
Aqua	May	05 2012; 12:55

P2172 L4-7: What do the authors mean with "elevation differences"? Large or small scale roughness? Some clarification regarding the influence of surface roughness on Ku band radar altimetry is needed here.
Here we mean large scale roughness, which is now clarified.

P2172 L17-18: But with the presence of a wet snow surface (like with the melt onset), shouldn't this favor rather narrow waveforms, because the wet layer on top prevents the radar signal

from penetration which also excludes scattering within the snow layer?
There is a transition from cold, dry snow which is transparent for the signal to a layer of water at the surface which has the characteristics of water. 'wet snow' has been replaced by 'snow with moderate temperature'.

P2172 L20: To which thickness is "thin ice" associated? It would be also interesting to know the fraction of frozen leads that are detected.
We agree, this would be very interesting. Unfortunately the thin ice thickness has so far not been derived directly from MODIS measurements in the visual range (in contrast to infrared but with a worse resolution). Another approach would be to estimate the time a lead exists to approximate the thickness of its ice cover. This is time intensive, not always possible (e.g. if the previous days are cloudy) and difficult to consider in the optimization. We define leads (including those covered by thin ice) as areas which are explicitly darker (less reflective) in the visual range than the surrounding ice. We expect the latest state of ice considered as a lead to be light nilas (5-10cm) but can not rule out the inclusion of grey ice (10-15cm).

Figure 2 slightly addresses the issue of frozen lead detection but it is at this point not possible to us to give a precise answer to this questions.

Depending on the freeboard of the "thin ice", this could introduce a positive bias in the sea-surface hight and hence a negative bias in freeboard.
We agree. This is mentioned at P2170L26-28

P2175 L14: Of which dimension is THETA?
This depends on the used classifier. It is now clarified:
'THETA consists of one threshold for each parameter used for the respective classifier.'

P2176 L5: Can the authors provide some more information (reference) about the Nelder-Mead simplex algorithm?
Done

P2175-P2176 Section 2.4: A structure chart of the analysis might improve the description of this method.
Included

P2176 L17-19: Does it mean that if there is a delay of +/-1 hour, measurements are discarded?
They are from the ground truth, yes. This is done in order to reduce the influences of ice drift and opening/closing of leads.

Figure 2a: What causes the gap between 71.3 and 71.5N ? Is it ambiguous regarding lead or

ice classification? On the image it clearly looks like ice only.
Yes it is ambiguous regarding lead or ice. After zooming in one can see a gray area slightly darker than the surrounding area crossing the track from the South-west to the North-east. Those areas are generally excluded as they might consist of narrow leads which are blurred due to the resolution of MODIS. It is part of our effort to exclude even narrow leads.

Figure 2: The color bar needs a label and units.
Corrected (it is unitless)

P2178 L26: But on the other hand the South Eastern Laptev Sea shows almost no leads which reveals the fast ice area in this region quite well!
We missed that, thank you.

Figure 5 and 6: Can the authors add that the data gap north of Canada is caused by the interferometric mode ("Wingham Box")?
Done

P2182 L5: Can the authors specify those "deficiencies"?
The following has been added:
'which might be caused by ice drift and opening/closing of leads between the data acquisition, collocation and unnoticed narrow leads'

P2182 L7: I think "effects" needs to be replaced by "affects".
corrected

P2183 Section 4.2: How do off-nadir leads affect the optimization? Are they completely excluded since CS-2 measurements with a mixture of classes in the footprint are discarded?
They are excluded as far as possible. Narrow off-nadir leads might still occur in the 'ice' class of the ground truth (due to the limited resolution of MODIS). This is now clearly stated. Narrow leads in the 'ice' class (nadir and off-nadir) shift the optimized threshold in the direction of the 'lead' class (i.e. increase it in the case of MAX and PP). We expect this shift to be small due to the limited amount of this leads (narrow leads are frequent compared to wide leads but still rare compared to ice). This shifts are further of minor importance as it only concerns the weight - threshold relation and not the more important threshold/parameter - performance relation.

P2185 Section 4.4: Can it be that the MAX threshold is optimized only for "large" leads due to the rejection of measurements with mixtures of both classes within the CS-2 footprint as well as the limited resolution of MODIS which was used as a reference in the optimization analysis?

This could also cause the small number of apparent lead widths of 300m.
This is another possible explanation and has been added.

Referee 3

However, I think the authors are too ambitious in labeling the false detection of leads with the MODIS data. The 250 m spatial resolution of MODIS is far too coarse to capture small leads (which are still seen by CryoSat-2) and have been found to be the most common type of leads in high resolution submarine sonar data. Based on the available data set, it would seem possible that only the true lead detection for large leads can be reliably stated, there is too much uncertainty in stating that false detections are present in other methods given the inability of MODIS to resolve small leads. This scaling back still provides worthwhile results, but better bounds the data within the lead detection ability of the control data set.
We agree. This problem is now clearly stated and addressed by an analysis of the SSH fluctuations.

'Note that this method is limited by the resolution of MODIS. CryoSat-2 measurements which look like originating from ice in MODIS scenes can actually contain small amounts of leads. See section 4.2 for a discussion of this circumstance.'

'It has been shown that leads which cover only a small fraction of a radar altimeter footprint can dominate the signal, due to the high amplitude of specular returns (Drinkwater, 1991). Therefore CS-2 detects leads which are simply not visible for MODIS despite its higher resolution. The fraction of this leads in the ice class of the ground truth cannot be quantified by our approach. Those narrow leads either cover the nadir point or not, while leads covering the whole footprint ('True Leads') do for sure. Therefore one could expect True Lead measurements to ensure a higher quality (see section 4.3) for the derivation of the SSH.

This expectation is supported by the smaller spread of the SSH estimate based on the MAX_1 compared to the PP_1 with nearly the same amount of lead detections (True + False Leads, Table 1). This advantage is on the other hand certifying that narrow, unnoticed leads in the ice class do not reverse the ROC analysis.'

In addition the focus on the False Lead Rate has been reduced, e.g. by replacing its mention it in the abstract and conclusion.

Another point that needs clarification is specification of the angles which are considered off-nadir (and therefore not used) in the classification scheme and control data sets.
As we use only SARM data we can not derive an angle between nadir and the main backscattering surface (lead) from CS-2 data like it has been done by Armitage and Davidson (2014). This is only possible if the SARInM is used (e.g. in the Wingham Box). If MODIS shows a lead within the CS-2 footprint (approximated by 300 X 1500 m) which does not cover it entirely, the measurement is discarded from the ground truth (and therefore from the optimization and control set). This is independent of the position of the lead (nadir/off-nadir). So there is no specific angle which separates nadir from off-nadir leads in the analysis. If taking about off-nadir leads we mean those leads which do not cover the nadir point but this is a purely linguistic definition.

Off-nadir data are always going to be present in the data so long as the pulse shape is broader than the transmit pulse, so this needs to be considered.
This is true, even though we are mainly interested in those cases where the off-nadir data is dominating the waveform as those can introduce a SSH bias.

Both of these main points tie in to a suggestion made by another reviewer, namely that the retrieval of surface elevation from identified lead points could be used to determine the impact on SSH determination. In my opinion, this would greatly enhance the results of the study.
We agree and added such analysis.

Other comments

2171, 14-16: What wavelengths are the MODIS bands? It would be helpful to have this in the text for those unfamiliar with MODIS.
Added

Section 2.1: It should be noted that the spatial resolution is quite different between CryoSat-2 and MODIS for the comparison. The CryoSat-2 footprint is not constant, and is largely a function of surface roughness. A rough surface will have a rectangular footprint size of 380 m x 1650 m, while a coherent scattering return from a smooth surface can still dominate the return even when the area is small, see Drinkwater 1991.
We agree and this is now noted. Besides the roughness (which has a stonger influence on the pulse limited (across track) part) also the satellite altitude is influencing the footprint size. We used the approximation of 300 X 1500m as given in Wingham et al. (2006).

Also one distinction to make is that the leads detected can be off-nadir, these need not be

labeled as false detections, but they should be accounted for if used in the retrieval of surface elevation.
As mentioned above, off-nadir leads are excluded (thereby not being labeled at all) if visible in the MODIS data. We derived the surface elevation without accounting for them and deduce from the lower SSH fluctuations of MAX_1 that the amount of off-nadir leads might be smaller. We are not able to say whether the more stable SSH is caused by less lead detections of off-nadir leads or of less actual ice measurements (or a combination of both). However we are confident, that the MAX_1 increases the quality of SSH estimate compared to other classifiers.

2171, 15-18: It is a bit ambiguous how leads are identified in the MODIS imagery, in particular, are only nadir leads considered, if not, how far off-nadir can a lead be?
Only leads covering the whole footprint (including the nadir point) are considered.

2172, 5-7: I am not sure what is meant by "favoring surface scattering instead of reflection"? This is just a confusion over word choice, perhaps it is meant that more energy is scattered away from the receiver?
corrected

2172, 13-18: The reflection and transmission of energy between the air and ice/ snow layers is probably not important compared to the geometric factors which affect the angular dependence of the backscattered energy.
This might be the case. But on the other hand we only state that there 'might' be an influence and Willat et al. (2011) state that the dominant scattering surface is somewhere within the snow layer if it has a temperature close to the freezing point. In addition Laxon (1994a) Page 918 reads: '[...] changes in return waveform power depend both on small scale surface roughness and on changing electrical properties.'.

2174, 3-5: What specific angle is defined as being off-nadir? This is an important distinction to make.
As mentioned above there is no specific angle (neither here nor in Ricker et al. (2014)). This lines have been rewritten in order to account for the existence of narrow, unnoticed off-nadir leads.

2176, 5: A description of the Nelder-Mead simplex algorithm would be beneficial here, along with some rationale for the chosen parameters.
Has been added.

The amount of initial guesses (i.e. repetitions of minimization from different starting points) has been increased until the global minimum was consistently found in several test cases. In a next step the number has been doubled (i.e. the interval between two

starting points was divided by two).

Section 3.1: A point brought up by another reviewer is that the MODIS data may not be showing some of the smallest leads due to 250 m resolution of the imagery. Lead width statistics are available from submarine sonar data (e.g. McLaren, 1989; Wadhams, 1981; Wadhams and Horne, 1980) and show that most leads are < 20 m. Thus, it is difficult to say with confidence that leads detected by CryoSat-2 and not by MODIS are false detections. Only the TLR for the largest leads can be determined with such a data set.
This problem is now discussed and addressed by estimates of the SSH.

2180, 8: The CS-2 track is very much two dimensional, the pulse-limited across-track footprint size is not negligible and could impact the results.
At this point we only describe the differences between the apparent lead width (CS-2 lead crossings) and actual lead width (short dimension of a lead). We focus on the difference between measurements fields (like images/MODIS) and line measurements (CS-2). The two dimensional nature of the footprint and its implications (e.g. off-nadir leads) are noted at several occasions throughout the text.

2187, 20: Some caution is needed in the lower threshold value as this number will depend on factors such as the altitude and that the transmit power is stable over time.
This is a good point, the waveform amplitude is dependent on the altitude and transmitted power. Changes of the transmitted power can be expected to influence the received power proportionally. Based on the relative high stability of the transmitter, we expect it to have a minor influence. We further expect the amplitude to be reduced with altitude (alt) by 1/(alt*2)^2 = 0.25*1/alt^2 (circular wave propagation). A typical CS-2 altitude is 717 km and we consider 10 km a typical variation (based on the data of January 2011), 727^2 / 717^2 = 1.028. The amplitude variation due to 10km change in altitude is therefore around 3%. This is much less than the variability within the class of assured leads (ground truth class 'lead') and can therefore be neglected. The MAX parameter has shown its advantages in lead detection on the ground truth and on a four year record of SSH estimates despite this two factors.

Manuscript prepared for The Cryosphere Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 29 May 2015

Lead detection in Arctic sea ice from CryoSat-2: quality assessment, lead area fraction and width distribution

A. Wernecke and L. Kaleschke

Institute of Oceanography, University of Hamburg, Bundesstrasse 53,
20146 Hamburg, Germany

Correspondence to: A. Wernecke (andreas.wernecke@uni-hamburg.de)

Abstract

Leads cover only a small fraction of the Arctic sea ice but they have a dominant effect on the turbulent exchange between the ocean and the atmosphere. A supervised classification of CryoSat-2 measurements is performed by a ~~combination~~ comparison with visual MODIS scenes. For several parameters thresholds are optimized and tested in order to reproduce this prior classification. The maximum power of the waveform shows the best classification properties amongst them, including the Pulse Peakiness. ~~With the same correct lead detection rates as of published classifiers, the amount of ice being detected as lead can be reduced by up to 40~~The sea surface height is derived and its spread is clearly reduced for a classifier based on the maximum power compared to published ones. Lead area fraction estimates based on CryoSat-2 show a major fracturing event in the Beaufort Sea in 2013. The resulting Arctic wide lead width distribution follows a power law with an exponent of 2.47 ± 0.04 for the winter seasons from 2011 to 2014, confirming and complementing a regional study based on a high resolution SPOT image.

1 Introduction

Sea ice ~~modulates~~ affects all interaction between ocean and atmosphere, namely heat, mass and momentum transports in ice covered regions. It ~~is strongly reducing~~ strongly reduces most of these transports and thereby leaving these processes basically to openings in the ice. These openings, called leads, appear even in regions which are typically covered by thick ice, like the central Arctic. Shear and divergence in the ice cover create new leads (Miles and Barry, 1998). Those areas can exhibit huge temperature differences between cold air and relative warm water. The resulting heat loss causes fast formation of new ice. Even leads covered by thin ice show much higher heat fluxes than the surrounding thick ice (Maykut, 1978). The low albedo of leads promotes an energy flow in the opposite direction which increases the amount of absorbed insolation, resulting in a warming of the underlying water. Leads reduce the internal strength of the sea ice, enabling higher drifting

velocities (Rampal et al., 2009) and are expected to influence the atmospheric boundary layer chemistry (e.g. Moore et al., 2014).

Large scale satellite remote sensing studies of lead occurrences have been done based on visual and thermal imagers (e.g. Lindsay and Rothrock, 1995; Willmes and Heinemann, 2015). They are generally limited by the resolution of thermal infrared measurements of about one kilometer and by the influence of clouds. By using passive microwave data, Röhrs et al. (2012) avoided the requirement of free sky conditions but reduced the resolution even further to 6.25 km. A good agreement with CryoSat-2 (CS-2) and ASAR based estimates of the lead occurrence for leads wider than 3 km has been reported (Röhrs et al., 2012). CryoSat-2 based lead detection is expected to be a good complement to previous estimates as it combines an increased resolution of some hundred meters with a strong atmospheric independence. The quality of this approach has been assessed by Zygmuntowska et al. (2013) for airborne surveys and is topic of this study for CS-2 measurements.

~~Not only~~ Apart from the lead area~~but~~ , also the width distribution is important for the turbulent heat transport in ice covered regions. A convective boundary layer evolves over leads which increases in thickness towards the downwind side of the lead (Andreas et al., 1979). This boundary layer ~~is dampening~~ dampens the heat flux per lead area which is therefore higher for narrow leads than for wider ones. This has led to different lead-width dependent heat transfer formulations (e.g. Andreas and Murphy, 1986). Marcq and Weiss (2012) show that the turbulent heat flux over leads is up to 55 % higher if using a power-law distribution down to a lead width of 10 m instead of considering all leads as one large area of open water.

The Arctic sea ice extent declined substantially over the last decades (Serreze et al., 2007), while comparable studies for the ice thickness are rare and struggle with uncertainties (Lindsay and Schweiger, 2015). Ice thickness estimates based on upward looking sonars on submarines (e.g. Rothrock et al., 2008) or moorings (Proshutinsky et al., 2009) have a relatively sparse temporal and spatial coverage. Airborne and helicopter based thickness measurements ~~are utilizing~~ utilize the strong difference between the electromagnetic inductances of seawater and ice. They are of great value for regional studies and valida-

tion, but are ~~limited to individual campaigns~~ restricted by the limited number of conducted surveys (Haas et al., 2010; Renner et al., 2013, 2014; Maaß et al., 2015).

Sea ice thickness is retrieved from satellites by radiometry, i.e. the influence of the ice thickness, salinity and temperature on the emissivity and transmittance. Various passive thermal to microwave sensors have been used (AVHRR, MODIS, SSM/I, AMSR-E, MIRAS) (Yu and Rothrock, 1996; Singh et al., 2011; Martin et al., 2005; Kaleschke et al., 2012; Tian-Kunze et al., 2014). As the ice thickness information saturates for all these sensors at a certain level, this approach is only capable of relatively thin ice, typically well below one meter (e.g. Kaleschke et al., 2010).

Another approach utilizes ~~laser or radar~~ altimetry in order to derive the snow or ice freeboard, i.e. the elevation difference between the Sea Surface Height (SSH) and snow or ice surface, respectively. Laser signals only reach the snow surface, while radar altimeters are basically showing the snow-ice interface elevation. By considering the relevant densities and the snow thickness ~~the freeboard~~ those freeboards can be converted into ~~an ice thickness in~~ ice thickness by assuming hydrostatic equilibrium. Sea ice thickness has been derived from ~~radar altimetry in the~~ K_u band radar altimetry from the European Remote Sensing Satellites ERS-1 and ERS-2 as well as Envisat and CS-2 (Laxon et al., 2003; Giles et al., 2008; Laxon et al., 2013; Ricker et al., 2014). These radars are not restricted to clear sky conditions, but ~~a satisfying handle for the influence of a snow cover on the signal is still not found~~limited knowledge of the snow loading and the radar interaction with the snow layer currently limits the accuracy of altimeter derived sea ice thickness estimates (Willatt et al., 2011; Kwok, 2014). Advantages of the radar on CS-2, over earlier K_u band altimeters are the reduced footprint size and noise due to the synthesis of overlapping measurements, its orbit which allows a coverage up to 88° North and South and the potential of interferometric measurements (Wingham et al., 2006). In most parts of the Arctic the SARM is used except for most coastal areas and the so called 'Wingham Box' (80-85° N and 100-140° W) where the SAR Interferometric mode finds application.

~~A crucial information for each~~ The SSH is crucial for altimeter based ice thickness ~~retrieval is the SSH~~retrievals. For this reason the altimeter measurements are separated into those

from ice and those from leads (see Fig. 1 for examples from CS-2). The lead measurements are used to derive the SSH, which acts as reference for the freeboard. Leads covered by thin ice and falsely detected leads (i.e. thick ice) result in an overestimation of the SSH and therefore in a negative bias in the derived freeboard and thickness. If considering only a very few, assured lead measurements the statistical error increases (Armitage and Davidson, 2014). It is therefore of high interest to find a lead detection method which is very trustworthy and detects as many leads as possible.

In this study the quality of CS-2 based lead detection procedures is assessed by a comparison with MODIS measurements. Previously published classifiers are implemented and compared with newly derived ones in a Receiver Operating Characteristics (ROC) graph. The most promising one is subsequently used to derive the lead area fraction and the lead width distribution. Thereby this study attempts to close a gap of knowledge about the differences of lead detection procedures from CS-2 and makes suggestions for improvements, which has direct implications for sea ice thickness estimates.

2 Methods

2.1 The ground truth

In order to optimize and compare the performance of different classification routines, we choose a supervised classification approach. ~~The~~ Visual Moderate Resolution Imaging Spectroradiometer (MODIS) measurements can be used to distinguish between sea ice and water (Su et al., 2012). Two MODIS instruments are in operation on the NASA satellites Terra and Aqua. They cover the earth surface every 1 to 2 days and measure in 36 spectral bands from visual (used here) to infrared (Barnes et al., 1998). We identify land and cloud influences manually and are therefore able to rely only on the MODIS band 2 ~~swath product is used~~ (around 857 nm wavelength) level 1B reflectance as reference data~~as it has a sufficient~~. It has a resolution of 250 m and seems to be even more suited to identify leads than band 1 (not shown). Dark areas with sharp edges and linear shapes in the MODIS

images are interpreted as leads. CS-2 measurements from ~~this~~ these areas, recorded less than one hour before or after the MODIS acquisition, are manually labeled as lead. In the same way we identify CS-2 measurements of ice while all measurements with a mixture of both classes within the footprint are excluded from this study (see also Fig. 2a). ~~This~~ The CS-2 footprint is assumed to be 300 m in and 1500 m across flight direction. This ice/lead information is in the following considered as ground truth, regardless of possible mislabeling for example caused by unexpected high ice velocities.

The ground truth consists 722 lead and 5768 ice measurements. Note that this method is limited by the resolution of MODIS. CryoSat-2 measurements which look like originating from ice in MODIS scenes can actually contain small amounts of leads. See section 4.2 for a discussion of this circumstance. The ground truth is acquired from February to ~~end of April~~ the beginning of May in 2012 and 2013 from seven MODIS granules in the eastern Beaufort sea and north of the Canadian Arctic Archipelago. For this time of the year optical MODIS scenes are available and surface melting can be ruled out. Within this study we use CryoSat-2 Level 1b data with processor versions “SIR1SAR/4.0” and “SIR1SAR/4.1” (Baseline B). This two SAR mode versions are equivalent.

2.2 Relation to physical properties

~~When the emitted CS-2 signal reaches a rough surface the elevation differences result~~ Large scale roughness results in a spread in time of the received ~~signal. Especially if the roughness is close to the signal wavelength of 2.2in~~ K_u band it is favoring CS-2 signal as exposed parts of the surface are reached earlier than low lying parts. Additionally small scale roughness is promoting surface scattering instead of reflection, ~~resulting in a more uniform distribution in terms of angle dependency~~meaning that more energy is scattered in oblique directions. Therefore measurements of the same position from altering look angles are more similar for rough surfaces (Wingham et al., 2006). Additionally areas further away from the nadir point have a stronger contribution, leading to an emphasized signal following the first (nadir) peak for rough surfaces ~~. As~~ (Laxon, 1994a). Energy conservation conditions

a reduced maximal receivable signal if the emitted power is ~~more strongly~~ scattered in all directions ~~, the maximal receivable signal is reduced~~by a rough surface.

The characteristic impedance of the surface layer might also influence the signal amplitude (Laxon, 1994a). If the difference in impedance at 13.5 GHz of the uppermost layer and the air is small, there is less reflection and more transmission into the ice/snow. Within the medium it is partly absorbed and scattered by inhomogeneities, again leading to a spread of the signal with lower maximum values and a more homogeneous angular distribution. This process could for example be favored by a layer of ~~wet snow~~ snow with moderate temperature.

As leads are locally bound, the fetch is too small for bigger waves to evolve in the water. The thin ice cover, if present, is yet neither physically deformed nor covered with snow. Furthermore the microstructure of young ice is more compact than of older ice as most brine pockets are filled and less connections have evolved. Therefore leads can be characterized by their commonly flat surface with relatively high impedance difference to the air. The returns originating from leads are expected to be compressed in time with higher maximum values and stronger incidence angle dependency (specular returns).

In the SAR Mode, CS-2 returns from the same position on the ground are received from different look angles. After allocating them by using the Doppler Shift, they are combined to one waveform (Wingham et al., 2006) i.e. the returned power as function of time (see Fig. 1 for examples). The following waveform based parameters are used: Maximum Power, Pulse Peakiness, Leading Edge Width and Trailing Edge Width.

In addition to stacking the beams to one waveform, they are also integrated over time ~~separately. In this case~~ (summed) separately. A Gaussian distribution curve is used to approximate the returned energy ~~of the individual beams (the stack) with different incidence angles is achieved~~as function of beams (i.e. incidence angle). We use the Stack Standard Deviation and the Stack Kurtosis ~~as stack based parameters~~ parameters which are based on this curve.

2.3 Parameter definition

– The Maximum Power (MAX) is the highest recorded power of the calibrated waveform in Watts.

– The Pulse Peakiness (PP) has been established by Laxon (1994b) and is defined as the MAX divided by the accumulated power (P^{WF}) of all bins constituting the waveform:

$$\mathrm{PP} = \frac{\max\left(P^{\mathrm{WF}}\right)}{\sum_{i=1}^{128} P_i^{\mathrm{WF}}}, \tag{1}$$

which is the same definition as used by Armitage and Davidson (2014), while the values of Laxon et al. (2013) are divided by 100 and those of Ricker et al. (2014) by 128 for consistency.

– The Left and Right Pulse Peakiness (PPL and PPR) from Ricker et al. (2014) for Baseline B data are defined as (Robert Ricker, personal communication, January 2015):

$$\mathrm{PPL} = \frac{15 \cdot \max\left(P^{\mathrm{WF}}\right)}{\sum_{i=i\max-6}^{i\max-2} P_i^{\mathrm{WF}}}, \tag{2}$$

$$\mathrm{PPR} = \frac{15 \cdot \max\left(P^{\mathrm{WF}}\right)}{\sum_{i=i\max+2}^{i\max+6} P_i^{\mathrm{WF}}}, \tag{3}$$

where $i$max is the index of the maximal value of the waveform. The PPL and PPR ~~are~~ have been proposed in order to reject off-nadir leads~~which are,~~ , the influence of which can not be quantified based on our methodology (see Sect. 2.1)~~, not expected to appear in our data~~. Therefore the PPL and PPR are not fully included in this study. However, they are defined as we use the classifier of Ricker et al. (2014) for comparisons.

– The Leading Edge Width (LEW) is defined as the width between 1 and 99 % of the amplitude of a Gaussian fit to the leading edge of the waveform. The fitted area starts at the first bin reaching one percent of the Maximum Power and ends at the second bin, following the first peak. The first peak is the first local maximum reaching at least 50 % of the Maximum Power. To avoid bimodal waveforms, we completely exclude measurements with a first peak smaller than 80 % of the Maximum Power from this study. Similar fits and constrains are used by Kurtz et al. (2014).

– The Trailing Edge Width (TEW) is defined as the width between 99 and 1 % of the amplitude of an exponentially decaying fit to the trailing edge of the waveform. The fitted area starts at the position of the Maximum Power and ends at the last bin (e.g. Legresy et al., 2005).

– The Stack Standard Deviation (SSD) is the standard deviation (SD) a random variable would have, if the ~~Gaussian fit to the integrated power (i.e. energy)~~ mentioned

Gaussian approximation of the energy as function of beams would be its Probability Density Function. It is therefore describing its width and not the SD of the energy values itself. The used unit is "Beams" which can also be expressed in terms of angles. Due to the more specular characteristics of leads, the spread of power with incidence angle is expected to narrow and so is the SSD (Wingham et al., 2006).

– The Stack Kurtosis (SK) is also obtained from the Gaussian fit to approximation of the energy as function of beams. In contrast to the SSD it is not derived as a fitting variable but from the function, evaluated at the positions of the beams (Veit Helm, personal communication, June 2014; Wingham et al., 2006). While all Gaussian functions have in general a kurtosis of three, discrete points of them do not have if the Gaussian is badly represented. This is the case for very narrow (badly resolved) and extremely wide functions (only partly covered). The Kurtosis is a measure of the peakedness which is expected to be high for leads.

2.4 Threshold optimization

Threshold based classifications are widely used to identify leads from K_u band altimeters. We use a repeated random sub-sampling validation Cross-Validation technique to derive and test thresholds (Θ). This (interested readers are referred to chapter 9 in Duda et al. (2001)). Θ consists of one threshold for each parameter used for the respective classifier. The Cross-Validation involves a random separation of the samples into a training and a testing subset, each of which consist of 50 % of all samples. From the training subset we derive Θ by using Eq. 4 and apply it to the testing set to investigate its performance. The random assignment into subsets and the testing of the newly derived Θ is repeated 200 times for each classifier to get an overall performance and an estimation of its spread. These steps are illustrated in Fig. 3.

As mentioned in Sect. 1 there are different applications for lead detection algorithms also resulting in different demands on its characteristics. One plausible aim is to reduce the total amount of false detections to a minimum. But one might also be interested in

a more conservative lead detection by reducing the amount of ice being detected as lead (False Leads) at the cost of less correctly detected leads (True Leads). A more conservative detection might be used for a freeboard retrieval as False Leads might result in a bias while high True Lead Rates are not always of high importance.

To take these different demands into account we include a weighting factor w in the cost function.

$$\text{cost}(\Theta) = w \cdot \text{False\_Ice}(\Theta) + \text{False\_Leads}(\Theta), \tag{4}$$

where False_Ice represents the amount of lead samples classified as ice. Θ is derived by minimizing the cost function on the training subset using the Nelder–Mead simplex algorithm (Nelder and Mead, 1965) with up to 400 initial guesses to find the global minimum. ~~For $0 < w < 1$ False Leads are primarily reduced~~The Nelder–Mead method is an unconstrained direct search algorithm for multidimensional minimization. This optimization reduces primarily False Leads for $0 < w < 1$, while for $w = 1$ the total amount of false classifications (False Ice + False Leads) is minimized. We use the parameter acronym with the weight as index to point at the corresponding one dimensional classifier.

This methodology is applied to all single parameters and all possible pairs of them~~, in~~. In the latter case ~~forming two dimensional feature spaces~~Θ is derived as the combination of both thresholds with the smallest value of the cost function.

3 Results

3.1 Classification performance

In Fig. 2 the CS-2 track is essentially crossing three wider leads, two of which are brighter at the northern side. This indicates that they are covered by ice on this side, while the southern side might exhibit open water. The third wider lead around 71.2° N and a thinner one at 71.75° N seem both to be completely covered by thin ice. The manual classification in Fig. 2a only visualizes the methodology as the time difference is larger than one hour

and this scene is therefore not part of the ground truth. Gaps in the track occur when the MODIS information of CS-2 footprints cannot be assigned unambiguously to leads or ice. In Fig. 2b the $MAX_{0.5}$ ~~classifier~~ detects only some leads and those only partly. It shows an accentuation on newly formed parts of the leads. The MAX_1 ~~classifier and the one~~ and the classifier developed by Ricker et al. (2014) (from now on called RI14) show strong similarities as both detect all relevant leads while lead detections are very rare where the MODIS scene shows ice. However for wider leads they can show a mixture of ice and lead detections, which is in some cases stronger for the MAX_1 ~~classifier~~ and in other cases for the ~~classifier~~ RI14 (not shown). The classifier ~~from~~used by Laxon et al. (2013) (from now on called LX13) detects all visible leads without a significant amount of missing lead detections, but it also detects leads where no or only weak indications for them can be found in the MODIS scene.

Figure 4 shows a Receiver Operating Characteristics (ROC) graph of all tested classifiers. Each classifier is represented by one point in the graph, the position of which is defined by its True Lead Rate (TLR, the amount of correctly detected leads divided by the amount of tested lead samples) and False Lead Rate (FLR, the amount of ice measurements in the ground truth detected as lead divided by the amount of tested ice samples). The upper left corner corresponds to ideal classifiers and the principle diagonal represents random assignments. For each parameter and pairs of them, we use different weights resulting in different Θ and corresponding performances. For single parameter classifiers 15 different weights (0.001, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1, 2, 5, 10, 30, 100) are applied to capture the development of the performance (from the lower left corner to the upper right in Fig. 4) while $w = 0.001$, $w = 0.5$ and $w = 1$ are used in the two dimensional case. To follow the performance of e.g. MAX based classifiers one can start with small w, implying high values of Θ which is detecting only a few leads (lower left corner in Fig. 4). With increasing w and decreasing Θ the TLR increases in the beginning much faster than the FLR. At some point the amount of correct lead detections is mostly constant while a further lowering of Θ is mainly increasing the amount of ice measurements which are detected as lead. As relative performances are shown the classifier closest to the upper left corner is not necessarily the

"best" one but if one classifier is on the upper left side of another it can be considered as superior. Further remarks on ROC graphs are given by Fawcett (2006).

It is at this point not important how the thresholds (Θ) are derived but only the combination of its value and performance. This allows us to compare the classifiers found here with those other authors have developed, independent from the optimization routine.

Classifiers based on the Maximum Power (MAX) appear on the upper left side of all others on the whole range of Fig. 4. Only classifiers using two parameters including the MAX (black marker) reach the single feature MAX classifications and are in all cases very close to them. All other classifiers based on pairs also show very similar results to that classifier based on the single, more suited parameter within its pair.

The PP and LEW based classifiers show strong similarities and have the second-best combinations of True and False Lead Rates.

Figure 5a illustrates the spread within the runs in terms of the SD of the True and False Lead Rates. The differences between all shown one dimensional classifiers and the corresponding two dimensional ones are clearly smaller than the inherent fluctuations and are therefore considered as not significant. The classifiers based on the MAX are separated from the others by more than their SDs for small weights, while they are not for higher weights. However, Fig. 5b shows that the fluctuations occur mostly in the direction of neighboring weights, which shows that the performance of MAX based classifiers is even stronger separated from the others.

3.2 Sea Surface Height

The SSH is calculated as a second stage of accessing the quality of classifiers. To derive the SSH from leads is a popular application, to test the classifier behavior in this context is therefore a very practical approach. This is done statistically by investigating the stability of SSH estimates from different classifiers.

We use the threshold first maximum re-tracker (TFMRA, TL=0.4) (Helm et al., 2014) to calculate the delay time which is assumed to correspond the return from the main scattering surface. The range is corrected for atmospheric influences (ionosphere, wet and

dry troposphere, dynamic atmosphere and the inverse barometric effect) and tides (namely: ocean, long period, solid earth, polar and ocean loading tides) as provided in the CS-2 L1B data. The surface elevation of lead measurements is considered as SSH and the variance is calculated on a $10\,\mathrm{km} \times 10\,\mathrm{km}$ north polar stereographic grid. All SARM measurements from January to March of the years 2011 to 2014 are used.

The mean field represents the stable SSH signal which could be used as reference for ice thickness estimates. The variance around it is an indicator for its reliability. Differences of the variance between the classifiers are expected to be caused by the inclusion/exclusion of ice measurements and/or off-nadir leads. Figure 6 shows the variance distribution of selected classifiers. The MAX_1 shows in general the smallest variances, while the amount of grid cells with high values converges to zero for all classifiers in a similar way.

3.3 Spatial distribution

In the following sections we are using the MAX_1 ~~classifier~~ which has been derived by minimizing the total amount of false classifications and its results are therefore taken as the best representation of the overall lead occurrence.

Figure 7 shows the lead fraction in the Arctic region as derived from CS-2 by dividing the amount of detected lead measurements by the total amount of measurements from January to March 2011. The AMSR-E Arctic lead area fraction (Röhrs and Kaleschke, 2012; Röhrs et al., 2012) (downloaded in September 2014) is also shown, combined over the same period and brought to the same spatial resolution.

Lead detections from CS-2 are most common in the Baffin Bay, the Fram Strait region, the northern Barents Sea and the Kara Sea, as well as in the western Laptev and the Chukchi Sea, all with lead fractions up to around 15 % (Fig. 7a). The central Arctic including the area north of the Canadian Arctic Archipelago and the northern Canada Basin show low lead fractions of around 0–1.5 %. In the southern Beaufort sea and especially its shear zone next to the coast line, lead fraction values of up to 6 % occur.

A somewhat different picture of the lead fraction pattern emerges by using the AMSR-E Arctic lead area fraction from Röhrs et al. (2012) (Fig. 7b). In areas covered by both esti-

mates the CS-2 based one mostly appears to be higher than the AMSR-E based estimate. This is not the case in the south eastern Beaufort sea where the AMSR-E product shows values of 15 % and more while they reach from 1.5 to 5 % for the CS-2 based estimate. We observe reasonable agreements in the Fram Strait region, the East Siberian Sea and the Chukchi Sea. Increased values occur for both estimates near islands like Svalbard, Franz Josef Land, Severnaya Zemlya and the Wrangel Island. However there are big differences between the datasets in the Baffin Bay, the Fram Strait regions close to the ice edge, the northern Barents Sea and the Kara Sea where CS-2 consistently detects more leads than the AMSR-E lead area fraction indicates.

While a daily open ocean mask is provided for the AMSR-E product, we consider all areas north of 65° N for the CS-2 based estimates. The ice edge on the Atlantic side, as indicated by the AMSR-E mask agrees well with the transition of CS-2 lead fractions from zero to higher values.

By the end of February 2013 the whole Beaufort Sea was pervaded by leads. Favored by storms the ice started in mid February to move into the direction of the Bering Strait, causing a divergence in the pack ice. This is the reason for the opening of leads, beginning in the western part and propagating to the east. This process accelerated around 27 February after which all but the fast ice at the Canadian coast and the sea ice at the Canadian Arctic Archipelago was fractured. See also Beitsch et al. (2014) for further descriptions.

By comparing the CS-2 lead fractions from February and March 2013 (Fig. 8) the pattern of this fracture event are reproduced with a proper shape and amplitude. Most lead pattern can be observed in both months, in many cases slightly decreasing in amplitude towards March. However, while in February only in the western part of the Beaufort Sea noticeable amounts of leads are detected, the complete region shows 8 to 15 % lead coverage in March.

3.4 Apparent lead width

To investigate the lead width distribution we use a proxy which we call apparent lead width. The apparent lead width is the amount of consecutive MAX_1 lead detections multiplied by

the approximate distance between two positions of 300 m. It can be seen as a measure of the CS-2 track interval over a crossed lead or as the width of a lead how it appears in the one dimensional domain of the CS-2 track. If the lead orientation is orthogonal to the CS-2 track, the apparent lead width is our best estimate of the actual lead width. We do not allow any ice detection within a lead which will in case of false detections split a lead into smaller ones.

The apparent lead width distribution follows a power-law in winter months with an exponent of 2.47 for values of 600 m and more (Fig. 9). A quantity z is called power-law distributed if its Probability Density Function $p(z)$ satisfies:

$$p(z) \propto z^{-a}, \tag{5}$$

where a is the power-law exponent. It is derived following the approximation of Clauset et al. (2009) for discrete distributions with a simple adjustment for a step size of 300 m as shown in Eq. (6).

$$a \approx 1 + N_Z \left(\sum_{i=1}^{N_Z} \ln \frac{z_i}{z_{\min} - \frac{1}{2} \cdot 300\,\mathrm{m}} \right)^{-1} \tag{6}$$

For the calculation of the power-law exponent only apparent lead widths z_i with a width equal or higher than $z_{\min} = 900\,\mathrm{m}$ are considered with N_Z being the amount of them. A line representing a power-law with the calculated exponent is displayed in Fig. 9. It shows the validity of this approximation down to 600 m, as both lines show a strongly parallel development.

The interannual variability is small with exponents between 2.42 in 2013 and 2.52 in 2011 with a SD of 0.04 amongst all four years. Differences between January, February and March of the same year are even smaller while the exponent decreases towards spring and autumn. All calculated distributions are following a power-law for apparent lead widths of 600 m and more.

4 Discussion

4.1 Classification performances

Classifiers based on the MAX parameter show generally the best ratio between ~~correctly detected leads and ice being falsely detected as lead~~True and False Lead Rate. A classifier using MAX $> 2.58 \times 10^{-11}$ W as threshold (MAX_1) detects 68.18 % of all leads correctly while only 3.41 % of the tested ice measurements in the ground truth are detected as leads. The PP_1 ~~classifier~~ using 0.35 as threshold has a TLR of 64.66 % (instead of 68.14 %) and a FLR of 4.09 % (instead of 3.41 %). The differences are even stronger for higher thresholds of 1.22×10^{-10} W and 0.425, respectively ($MAX_{0.5}$ and $PP_{0.5}$, Table 1).

The performances of individual runs overlap only slightly for $w = 1$ and are well separated for $w = 0.5$. This shows that the performance improvement is significant. The increased fluctuation in the direction of neighboring weights in Fig. 5b is likely to be caused by a variability of the thresholds caused by the repeated optimization.

For airborne surveys with ~~an~~ a device very similar to SIRAL on CS-2, Zygmuntowska et al. (2013) also found the MAX parameter to have less false lead classifications than all other parameters. The best combination of parameters (MAX & TEW) with ~~an~~ a Bayesian classifier improves its rate only little from 6.5 to 6.2 %. Zygmuntowska et al. (2013) define the False Lead Classification Rate (FLCR) as percentage of all lead detections originating from sea ice. This is different to our False Lead Rate as we use the amount of true ice measurements as base. The FLCR calculated from the absolute values in Table 1 are 28.6 and 12.5 % for ~~the~~ MAX_1 and $MAX_{0.5}$~~classifier~~, respectively. One reason for higher error rates of CS-2 is the reduced resolution of 300 m $\times$ 1500 m in contrast to around 10 m $\times$ 50 m for the airborne device. Thereby it becomes much more likely that different surface types occur within one footprint. Further we have to allow for some temporal differences in the data acquisition and have to collocate the datasets, while for the airborne surveys optical images are taken simultaneously. Deficiencies of the ground truth which might be caused by ice drift and opening/closing of leads between the data acquisition, collocation and unnoticed narrow leads increase the error rates which might therefore be overestimated.

However, as this effects all parameters it is likely that the influence on the comparison is small.

Compared to their MAX classifier, the PP classifier of Zygmuntowska et al. (2013) is detecting more leads from both, ice and lead measurements. This is directly connected to applied thresholds and is not a parameter property. For a solid decision which parameter is suited best for lead detection it is necessary to vary the thresholds. We are able to propose the use of the MAX parameter instead of the PP.

Three classifiers developed by other authors are included in this study. With the same amount of False Lead detectionsLeads, the True Lead Rates can be increased for our dataset from 9 to $\sim 13\,\%$ (Röhrs et al., 2012), from 83 to $\sim 89\,\%$ (LX13) or from 61 to $\sim 79\,\%$ if an optimized threshold of the MAX parameter (RI14) if a MAX based classifier is used instead . On the other hand the False Lead Rate can be reduced for a constant True Lead Rate from 13 to $\sim 7 (-38)$ or from 5 to $\sim 3 (-40)$for the and classifier, respectively (Fig.(Fig. 4).

So far we defined a lead detection as a measurement reaching both thresholds , but the same procedure has been performed by taking it as sufficient to reach only one of them. This results in more demanding thresholds but like before in The shown classifiers using two parameters detect a lead if both thresholds are reached. This logical 'and' criterion is now replaced by an 'or'. A classifier based on the MAX and the PP could for example define a measurement as a lead if its MAX value is above $10^{-}11$ W or if its PP value is above 0.3 (one of those is now sufficient). This is influencing the amount of False Ice and False Lead detections (i.e. the cost function). As a result our example has higher thresholds than it would have for the same weight and parameters using the 'and' criterion. Performing the same test as before but now using the 'or' criterion for all pairs of parameters brought no improvement of the classification (not shown).

There seems to be no benefit in including a second feature, because of which the optimization shifts the second threshold to some level where it does not harm. The lack of improvement

The fact that combining two parameters seems to have no benefit at all indicates that the parameters are basically all utilizing the same physical information and that the instrument and fading noises have either an correlated influence on all parameters or the influence is not significant at all. As some of the parameters are derived in a very different way ~~, firstly by stacking different beams and secondly by integrating them over time separately,~~ (e.g. waveform and stack based ones) we do not expect the noise to affect them ~~in the same way. We further conclude that the spread of the parameters within one class is predominantly caused by variations of surface properties which produces a correlation of all parameters~~equally. We conclude that noise plays probably a minor role for the classification errors.

4.2 Narrow leads and Sea Surface Height

It has been shown that leads which cover only a small fraction of a radar altimeter footprint can dominate the signal, due to the high amplitude of specular returns (Drinkwater, 1991). Therefore CS-2 detects leads which are simply not visible for MODIS despite its higher resolution. The fraction of this leads in the ice class of the ground truth cannot be quantified by our approach. Those narrow leads either cover the nadir point or not, while leads covering the whole footprint ('True Leads') do for sure. Therefore one could expect True Lead measurements to ensure a higher quality (see section 4.3) for the derivation of the SSH.

4.3 ~~Off-nadir leads~~

This expectation is supported by the smaller spread of the SSH estimate based on the MAX_1 compared to the PP_1 with nearly the same amount of lead detections (True + False Leads, Table 1). This advantage is on the other hand certifying that narrow, unnoticed leads in the ice class do not reverse the ROC analysis.

~~Due to the high amplitude of returns from leads it is possible that~~

4.3 Off-nadir leads

As mentioned before, leads which are not directly in nadir direction ~~are dominating~~ can dominate the signal. As this can cause a bias in elevation estimates Ricker et al. (2014) introduced the Left and Right Pulse Peakiness to avoid off-nadir leads. It has further been shown that it is, to some extent, possible to reduce the influence of off-nadir leads by increasing the Pulse Peakiness threshold of a single parameter classifier (Armitage and Davidson, 2014). This is done at the cost of discarding $\sim 60\,\%$ of the lead detections and thereby increasing the statistical error. The underlying process allowing for this reduction is the influence of the surface orientation towards the sensor on the maximum return. The relative orientation, favoring high maximum values the most, is expected to be found close to the nadir point. The further away from this point the main scattering surface (i.e. the lead) is, the more power is reflected away from the sensor instead of back towards it. This process is influencing the MAX value in the first place which has then implications for the PP (Armitage and Davidson, 2014). Therefore it is reasonable to assume that the influence of off-nadir leads is also reduced for high MAX thresholds, potentially even stronger than for the PP as the process causing this reduction has a more direct impact on it. This assumption is supported by the reduced SSH variance of the MAX_1 even though we can not say whether this reduction is caused by the elimination of off-nadir leads or incorrectly classified ice measurements (or a combination of both).

4.4 Spatial distribution

The CS-2 lead fraction shows a reasonable spatial distribution. It is small in the central Arctic and north of the Canadian Arctic Archipelago which are typical regions of thick multi-year ice. It shows high values in regions of high drifting velocities or known to favor the development of polynyas like the Fram Strait, the western Laptev Sea and the Chukchi Sea. The lead fractions are also increasing around most islands and coasts which introduce shear between the land fast ice and the drifting pack ice. Small lead fractions in the eastern Laptev sea and the western parts of the East Siberian Sea could indicate the presence of

large amounts of land fast ice. The absolute lead fraction values tend to be higher but are mostly in agreement with those of Lindsay and Rothrock (1995). They found lead fractions of 2 to 3 % for the central arctic and 6 to 9 % in the peripheral seas in the winter using the Advanced Very High Resolution Radiometer (AVHRR).

In nearly all regions the CS-2 lead fraction exceeds the AMSR-E Arctic lead area fraction from Röhrs et al. (2012) (Fig. 7). While the AMSR-E product only detects most leads with a width of 3 km and more, a width of at least some hundred meters is sufficient for detection by CS-2. As shown in Sect. 3.4 the apparent lead width is following a power law on the scale of kilometers, implying that measurements from narrow leads are largely outnumbering those from wider leads. In contrast to the CS-2 lead fraction, the AMSR-E product is additionally not including very large regions of thin ice like huge polynyas as a spatial high-pass filter is used.

The ice edge towards the North Atlantic is captured by both approaches quite similar. ~~We~~ In Fig. 7a we expect the ice edge to be at the interface between areas of no lead detections around the Norwegian and central Barents Sea and neighboring areas of higher lead fractions~~in Fig. 7a~~. This allows the inference that the MAX_1 ~~classifier~~ detects no leads over the open ocean. For this reason grid cells at the very ice edge are likely to underestimate the lead fraction relative to the ice covered part of the cell.

While the AMSR-E lead fraction drops relative consistently down to values around 2–3 % within a belt of around 200 km from the ice edge, CS-2 based estimates show much higher values of around 12 % in this areas. The high values in the marginal ice zone are reasonable as this area is likely to be fractured due to the influence of ocean waves. Especially in the Baffin Bay, the northern Barents Sea and the Kara Sea high rates of new ice formation can occur in winter which is in good agreement with high CS-2 lead fractions of these regions. The general reasonable distribution and its alternation increase our confidence in our lead detection algorithm.

4.5 Apparent lead width

Compared to the power-law, the found number of apparent lead width of 300 m is smaller than expected. This is a typical feature on the lower bound of the resolution as leads of this size are not always covered by a single measurement but partially by more, not necessarily leading to a detection. This is intensified by the elongated footprint of CS-2 as small leads may only be detected if they cover most of the width of the footprint. The MAX_1 is optimized mainly on leads wider than a single measurement which could also cause the relative small number of apparent lead width of 300 m. Therefore it is likely that the bend on the lower bound of the distribution in Fig. 9 is ~~induced by the measurement characteristics and not by the actual~~ an artifact and not a valid part of the lead distribution.

Marcq and Weiss (2012) have found a power-law exponent similar to our between 2.1 and 2.6 for scales from 20 m to 2 km by analyzing a single SPOT image with a resolution of 10 m. In two submarine based surveys, power-laws with exponents of 2 and 2.29 were found for the regions from the Fram Strait to the North Pole and the Davis Strait, respectively (Wadhams, 1981; Wadhams et al., 1985). In both cases resolutions of about 5 m are present and the power-law holds for the range from 50 to 1000 m. The examination of submarine and mooring data by Kwok et al. (2009) also indicates a strong accumulation of lead widths down to 5 m but the distribution has not been analyzed. For the central Arctic, a study of Lindsay and Rothrock (1995) also states a power-law distribution, but with a mean exponent of 1.6 for scales from one to around 50 km. It is based on thermal to near visible infrared measurements from the AVHRR, which is despite its resolution of one kilometer expected to detect leads with a minimum width below this size. It has been discussed whether the lead width distribution might be scale dependent (Lindsay and Rothrock, 1995; Marcq and Weiss, 2012) which seems not to be the case, as we found a stable power-law behavior on scales partly covering those of all other studies.

The results of Lindsay and Rothrock (1995) are contradictory to ours as we found a higher power-law exponent, implying a higher fraction of narrow leads. One explanation would be the relative coarse resolution of the AVHRR in combination with its high sensitivity to

leads. This could cause leads to appear wider then they are, as well as several narrow ones to appear as one wide lead, resulting in a less steep apparent lead width distribution. Comparisons with MODIS images indicate that the classifier used in this study switches in some cases between lead and ice detections over refrozen leads. This could result in an overestimation of the power-low exponent. The estimates might also have a different tolerance of refrozen leads, while both include at least the early stages of freezing. The size of leads is often growing with time as the surrounding ice floes keep drifting apart, meaning that estimates which include older leads are also likely to show less steep apparent lead width distributions.

Another reason could be an actual shift in the distribution between the periods from 1989 to 1995 and 2011 to 2014. This would be consistent with the results of Marcq and Weiss (2012) but would not explain the differences to those studies by Wadhams (1981) and Wadhams et al. (1985). However, this shift could be driven by observed changes in the amount of perennial ice, the ice thickness and drifting velocities (Nghiem et al., 2007; Haas et al., 2008; Rampal et al., 2009). Rampal et al. are further linking a found increase in winter strain rates between 1978 and 2007 to a weakening in mechanical strength of the ice and increased fracturing. We found no sign for a trend of the power-law exponent within the four years of CS-2 data.

4.6 Implications of apparent lead width distribution

As most leads are not crossed orthogonally, the apparent lead width is typically larger than the actual width of the lead. A transformation ~~from apparent to actual lead width~~ to the latter is not possible without profound knowledge of the sensitivity of lead detections and requires assumptions about the shape and orientation of leads. This is impeded by a nonuniform distribution of lead orientation (Bröhan and Kaleschke, 2014). For most applications it is not necessary to perform this transformation as this is the way leads appear to anything moving along sea ice, including the wind acting on the ocean surface.

The apparent lead width distribution is showing a strong intensification towards smaller lead widths. The area contribution of leads having the width z is $z \cdot p(z) \propto z^{-2.47+1}$, which

is still relatively fast decreasing with increasing width. This indicates that every lead area estimate which is not capable of detecting narrow leads is very likely to underestimate the total lead area. For a parametrization of lead area estimates it is of high interest to know down to which bound the power law behavior holds. This defines not only the mean lead width but also the fraction of lead area which is not captured by the estimate.

5 Conclusions

This study presented the potentials of several parameters and combinations of them to distinguish between CryoSat-2 measurements from leads and those from ice. They have been tested by deriving thresholds and analyzing their capabilities of reproducing a prior classification. The combination of parameters, even though common practice, has not shown any advantage for threshold based classifications. Using the maximum value of the waveform has in all cases shown better results than any other tested parameter, including the Pulse Peakiness. Compared to the classifier ~~of~~used by Laxon et al. (2013) a threshold of ~~1.22×10^{-10}~~2.58×10^{-11} W on the MAX detected only ~~half the amount of true leads but was able to reduce the percentage~~ 68 instead of 83 % of ensured lead measurements but showed a much more stable SSH estimate by reducing the amount of ice being detected as lead ~~from around 13 to less than 1~~and/or off-nadir leads. A solid lead detection, which ensures that nearly all lead classifications are actually originating from leads is ~~the requirement for~~ facilitating a precise, unbiased freeboard retrieval. It thereby helps to improve ice thickness estimates, which is one of the major aims of the CryoSat-2 mission.

~~A lower~~ The threshold of 2.58×10^{-11} W was further used as the best representation of the overall lead occurrence. It showed reasonable spatial distributions with relatively high lead fractions of around 12 % in the marginal ice zone. The apparent lead width was derived from the amount of consecutive lead detections. Its distributions is following a power-law with exponent of 2.47 ± 0.04 which implies a concentration of both, amount and area contribution at small lead widths. Embedding this work into those of others, a scale independent lead width distribution from 20 m to 50 km is likely. The implications for a parametrization

of low resolution lead area estimates were addressed and its dependency on the lower bound of the found distribution emphasized. The turbulent heat transport over ice covered regions is known to be strongly lead width dependent on small scales. The found distribution is suggesting that the work of Marcq and Weiss (2012), based on a single SPOT scene, can be generalized. This implies a much higher heat transport per lead area than it would be obtained by wide leads. In this manner the presented findings can help to improve the parameterization of this fundamental process in coupled ocean–ice–atmosphere models.

The Supplement related to this article is available online at doi:10.5194/tcd-0-1-2015-supplement.

Acknowledgements. This work is supported by the German Federal Ministry of Education and Research (FKZ: 01LP1126A) and in parts through the Cluster of Excellence "CliSAP" (EXC177), University of Hamburg, funded through the German Science Foundation (DFG). CryoSat-2 data are provided by the European Space Agency (ESA). Additionally we would like to acknowledge the use of Rapid Response imagery from the Land Atmosphere Near-real time Capability for EOS (LANCE) system operated by the NASA/GSFC/Earth Science Data and Information System (ESDIS) with funding provided by NASA/HQ. The AMSR-E Arctic lead area fraction is provided by the Integrated Climate Data Center (ICDC, http://icdc.zmaw.de/), University of Hamburg, Hamburg, Germany. We would like to thank Veit Helm, Stefan Hendricks and Robert Ricker for the kind and fruitful discussions.

References

Andreas, E. L., Paulson, C. A., William, R. M., Lindsay, R. W., and Businger, J. A.: The turbulent heat flux from Arctic leads, Boundary-Layer Meteorology, 17, 57–91, doi:10.1007/BF00121937, 1979.

Andreas, E. L. and Murphy, B.: Bulk transfer coefficients for heat and momentum over leads and polynyas, J. Phys. Oceanogr., 16, 1875–1883, 1986.

Armitage, T. W. and Davidson, M. W.: Using the interferometric capabilities of the ESA CryoSat-2 mission to improve the accuracy of sea ice freeboard retrievals, IEEE T. Geosci. Remote, 52, 529–536, 2014.

Barnes, W. L., Pagano, T. S., and Salomonson, V. V.: Prelaunch characteristics of the moderate resolution imaging spectroradiometer (MODIS) on EOS-AM1, Geoscience and Remote Sensing, IEEE Transactions on, 36, 1088–1100, doi:10.1109/36.700993, 1998.

Beitsch, A., Kaleschke, L., and Kern, S.: Investigating high-resolution AMSR2 sea ice concentrations during the February 2013 fracture event in the Beaufort Sea, Remote Sens., 6, 3841–3856, 2014.

Bröhan, D. and Kaleschke, L.: A nine-year climatology of Arctic sea ice lead orientation and frequency from AMSR-E, Remote Sens., 6, 1451–1475, 2014.

Clauset, A., Shalizi, C. R., and Newman, M. E.: Power-law distributions in empirical data, SIAM Rev., 51, 661–703, 2009.

Duda, R. O., Hart, P. E., and Stork, D.G.: Pattern classification, John Wiley & Sons, second edition., 2001.

Drinkwater, M. R.: K u band airborne radar altimeter observations of marginal sea ice during the 1984 Marginal Ice Zone Experiment, J. Geophys. Res.-Oceans (1978–2012), 96, 4555–4572, 1991.

Fawcett, T.: An introduction to ROC analysis, Pattern Recogn. Lett., 27, 861–874, 2006.

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.

Haas, C., Pfaffling, A., Hendricks, S., Rabenstein, L., Etienne, J.-L., and Rigor, I.: Reduced ice thickness in Arctic Transpolar Drift favors rapid ice retreat, Geophys. Res. Lett., 35, L17501, doi:10.1029/2008GL034457, 2008.

Haas, C., Hendricks, S., Eicken, H., and Herber, A.: Synoptic airborne thickness surveys reveal state of Arctic sea ice cover, Geophys. Res. Lett., 37, L09501, doi:10.1029/2010GL042652, 2010.

Helm, V., Humbert, A., and Miller, H.:Elevation and elevation change of Greenland and Antarctica derived from CryoSat-2, The Cryosphere,8, 1539–1559, doi:10.5194/tc-8-1539-2014, 2014.

Kaleschke, L., Maaß, N., Haas, C., Hendricks, S., Heygster, G., and Tonboe, R. T.: A sea-ice thickness retrieval model for 1.4 GHz radiometry and application to airborne measurements over low salinity sea-ice, The Cryosphere, 4, 583–592, doi:10.5194/tc-4-583-2010, 2010.

Kaleschke, L., Tian-Kunze, X., Maaß, N., Mäkynen, M., and Drusch, M.: Sea ice thickness retrieval from SMOS brightness temperatures during the Arctic freeze-up period, Geophys. Res. Lett., 39, L05501, doi:10.1029/2012GL050916, 2012.

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.: Simulated effects of a snow layer on retrieval of CryoSat-2 sea ice freeboard, Geophys. Res. Lett., 41, 5014–5020, 2014.

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

Laxon, S. W.: Sea ice altimeter processing scheme at the EODC, International Journal of Remote Sensing, 15, 915–924, doi:10.1080/01431169408954124, 1994a.

Laxon, S. W.: Sea ice extent mapping using the ERS-1 radar altimeter, EARSeL Adv. Remote Sens., 3, 112–116, ~~1994.~~ 1994b.

Laxon, S. W., Peacock, N., and Smith, D.: High interannual variability of sea ice thickness in the Arctic region, Nature, 425, 947–950, 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, 2013.

Legresy, B., Papa, F., Remy, F., Vinay, G., van den Bosch, M., and Zanife, O.-Z.: ENVISAT radar altimeter measurements over continental surfaces and ice caps using the ICE-2 retracking algorithm, Remote Sens. Environ., 95, 150–163, 2005.

Lindsay, R. and Rothrock, D.: Arctic sea ice leads from advanced very high resolution radiometer images, J. Geophys. Res.-Oceans, 100, 4533–4544, 1995.

Lindsay, R. and Schweiger, A.: Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations, The Cryosphere, 9, 269–283, doi:10.5194/tc-9-269-2015, 2015.

Maaß, N., Kaleschke, L., Tian-Kunze, X., Mäkynen, M., Drusch, M., Krumpen, T., Hendricks, S., Lensu, M., Haapala, J., and Haas, C.: Validation of SMOS sea ice thickness retrieval in the northern Baltic Sea, Tellus A, 67, 24617, doi:10.3402/tellusa.v67.24617, 2015.

Marcq, S. and Weiss, J.: Influence of sea ice lead-width distribution on turbulent heat transfer between the ocean and the atmosphere, The Cryosphere, 6, 143–156, doi:10.5194/tc-6-143-2012, 2012.

Martin, S., Drucker, R., Kwok, R., and Holt, B.: Improvements in the estimates of ice thickness and production in the Chukchi Sea polynyas derived from AMSR-E, Geophys. Res. Lett., 32, L05505, doi:10.1029/2004GL022013, 2005.

Maykut, G. A.: Energy exchange over young sea ice in the central Arctic, J. Geophys. Res.-Oceans, 83, 3646–3658, 1978.

Miles, M. W. and Barry, R. G.: A 5-year satellite climatology of winter sea ice leads in the western Arctic, J. Geophys. Res.-Oceans, 103, 21723–21734, 1998.

Moore, C. W., Obrist, D., Steffen, A., Staebler, R. M., Douglas, T. A., Richter, A., and Nghiem, S. V.: Convective forcing of mercury and ozone in the Arctic boundary layer induced by leads in sea ice, Nature, 506, 81–84, doi:10.1038/nature12924, 2014.

Nelder, J. A. and Mead, R.: A simplex method for function minimization, The computer journal, 7, 308–313, doi:10.1093/comjnl/7.4.308, 1965.

Nghiem, S., Rigor, I., Perovich, D., Clemente-Colon, P., Weatherly, J., and Neumann, G.: Rapid reduction of Arctic perennial sea ice, Geophys. Res. Lett., 34, L19504, doi:10.1029/2007GL031138, 2007.

Proshutinsky, A., Krishfield, R., Timmermans, M.-L., Toole, J., Carmack, E., McLaughlin, F., Williams, W. J., Zimmermann, S., Itoh, M., and Shimada, K.: Beaufort Gyre freshwater reservoir: state and variability from observations, J. Geophys. Res.-Oceans, 114, C00A10, doi:10.1029/2008JC005104, 2009.

Rampal, P., Weiss, J., and Marsan, D.: Positive trend in the mean speed and deformation rate of Arctic sea ice, 1979–2007, J. Geophys. Res.-Oceans, 114, C05013, doi:10.1029/2008JC005066, 2009.

Renner, A. H. H., Hendricks, S., Gerland, S., Beckers, J., Haas, C., and Krumpen, T.: Large-scale ice thickness distribution of first-year sea ice in spring and summer north of Svalbard, Ann. Glaciol., 54, 13–18, 2013.

Renner, A. H. H., Gerland, S., Haas, C., Spreen, G., Beckers, J. F., Hansen, E., Nicolaus, M., and Goodwin, H.: Evidence of Arctic sea ice thinning from direct observations, Geophys. Res. Lett., 41, 5029–5036, doi:10.1002/2014GL060369 2014.

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.

Röhrs, J. and Kaleschke, L.: An algorithm to detect sea ice leads by using AMSR-E passive microwave imagery, The Cryosphere, 6, 343–352, doi:10.5194/tc-6-343-2012, 2012.

Röhrs, J., Kaleschke, L., Bröhan, D., and Siligam, P. K.: Corrigendum to "An algorithm to detect sea ice leads by using AMSR-E passive microwave imagery" published in The Cryosphere, 6, 343–352, 2012, The Cryosphere, 6, 365–365, doi:10.5194/tc-6-365-2012, 2012.

Rothrock, D., Percival, D., and Wensnahan, M.: The decline in arctic sea-ice thickness: separating the spatial, annual, and interannual variability in a quarter century of submarine data, J. Geophys. Res.-Oceans, 113, C05003, doi:10.1029/2007JC004252, 2008.

Serreze, M. C., Holland, M. M., and Stroeve, J.: Perspectives on the Arctic's shrinking sea-ice cover, Science, 315, 1533–1536, 2007.

Singh, R. K., Oza, S. R., Vyas, N. K., and Sarkar, A.: Estimation of thin ice thickness from the advanced microwave scanning radiometer-EOS for coastal polynyas in the Chukchi and Beaufort Seas, IEEE T. Geosci. Remote, 49, 2993–2998, 2011.

Su, H., Wang, Y., and Yang, J.: Monitoring the spatiotemporal evolution of sea ice in the Bohai Sea in the 2009–2010 winter combining MODIS and meteorological data, Estuaries and coasts, 35, 281–291, doi:10.1007/s12237-011-9425-3, 2012.

Tian-Kunze, X., Kaleschke, L., Maaß, N., Mäkynen, M., Serra, N., Drusch, M., and Krumpen, T.: SMOS-derived thin sea ice thickness: algorithm baseline, product specifications and initial verification, The Cryosphere, 8, 997–1018, doi:10.5194/tc-8-997-2014, 2014.

Wadhams, P.: Sea-Ice Topography of the Arctic Ocean in the Region 70° W to 25° E, Philos. T. Roy. Soc. S.-A., 302, 45–85, 1981.

Wadhams, P., McLaren, A. S., and Weintraub, R.: Ice thickness distribution in Davis Strait in February from submarine sonar profiles, J. Geophys. Res.-Oceans, 90, 1069–1077, 1985.

Willatt, R., Laxon, S., Giles, K., Cullen, R., Haas, C., and Helm, V.: Ku-band radar penetration into snow cover on Arctic sea ice using airborne data, Ann. Glaciol., 52, 197–205, 2011.

Willmes, S. and Heinemann, G.: Pan-Arctic lead detection from MODIS thermal infrared imagery, Ann. Glaciol., 56, 29–37, 2015.

Wingham, D., Francis, C., Baker, S., Bouzinac, C., Brockley, D., Cullen, R., de Chateau-Thierry, P., Laxon, S., Mallow, U., Mavrocordatos, C., Phalippouc, L., Ratierb, G., Reyc, L., Rostand, F., Viaub,

P. and Wallisa, D. W.: CryoSat: A mission to determine the fluctuations in Earth's land and marine ice fields, Adv. Space Res., 37, 841–871, 2006.

Yu, Y. and Rothrock, D.: Thin ice thickness from satellite thermal imagery, J. Geophys. Res.-Oceans, 101, 25753–25766, 1996.

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. Selected classifier performance. The lower three classifiers are ~~derived~~ used by (from bottom to top): Ricker et al. (2014), Röhrs et al. (2012) and Laxon et al. (2013). TL: True Leads; FL: False Leads; TI: True Ice; FI: False Ice; TLR and FLR: True and False Lead Rates [%]; eTLR and eFLR: SDs of TLR and FLR within runs [%]. A list of all tested classifier performances is provided as Supplement.

Feat	w	Θ	TL	FL	TI	FI	TLR	FLR	eTLR	eFLR
MAX	10^{-3}	4.28×10^{-10}	10 336	226	576 597	61 841	14.32	0.04	2.07	0.05
MAX	0.5	1.22×10^{-10}	29 435	4220	572 634	42 711	40.80	0.73	3.26	0.24
MAX	1	2.58×10^{-11}	49 204	19 689	557 143	22 964	68.18	3.41	5.89	0.73
MAX	1	2.55×10^{-11}	48 808	19 580	557 305	23 307	67.68	3.39	5.61	0.72
TEW		200								
PP	0.5	0.425	22 677	7728	569 244	49 351	31.48	1.34	8.62	0.71
PP	1	0.35	46 602	23 623	553 307	25 468	64.66	4.09	4.94	0.66
PP	–	0.18	59 809	73 003	504 042	12 146	83.12	12.65	1.40	0.42
SSD		4								
MAX	–	6×10^{-10}	6576	00	576 811	65 613	9.11	0.00	1.12	0.00
PP	–	0.3125	43 875	28 599	548 145	28 381	60.72	4.96	1.74	0.27
SSD		4								
SK		40								
PPL		40								
PPR		30								

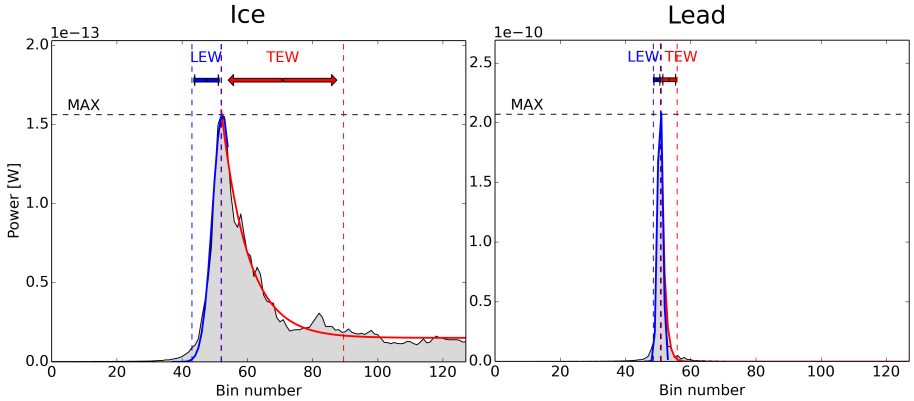


Figure 1. Example CryoSat-2 waveforms from ice (left panel) and a lead (right panel). The definition of the Leading Edge Width (LEW), Trailing Edge Width (TEW) and Maximum Power (MAX) are illustrated while the Pulse Peakiness (PP) is inversely proportional to the gray areas of normalized waveforms. The Bin number can be converted into delay time. Note the different scaling factors of the y axis ($\times 10^{-13}$ and $\times 10^{-10}$ for the ice and lead example, respectively).

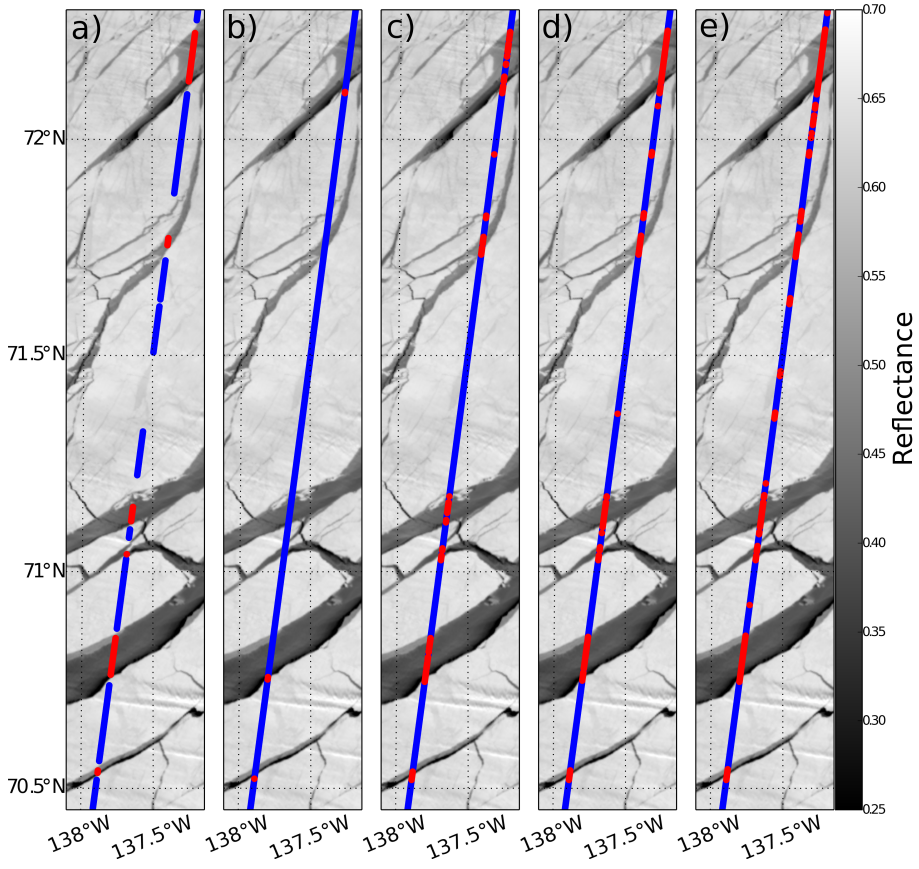


Figure 2. MODIS band 2 scene from 6 March 2013 in the southern Beaufort sea combined with a CS-2 track taken 83 min later on. The CS-2 samples have been classified as lead (red) and ice (blue) manually **(a)** or by ~~the application of thresholds from the **(b)**~~ $MAX_{0.5}$ **(b)**, ~~**(c)**~~ MAX_1 **(c)**, RI14 **(d)** ~~-~~and LX13 **(e)** ~~classifier~~. The classifier from Röhrs et al. (2012) detects no leads within this section.

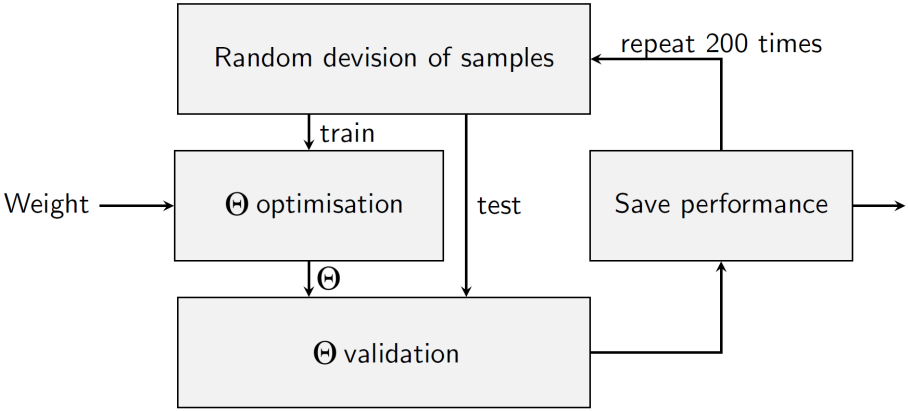


Figure 3. Flow chart of the used Cross-Validation scheme.

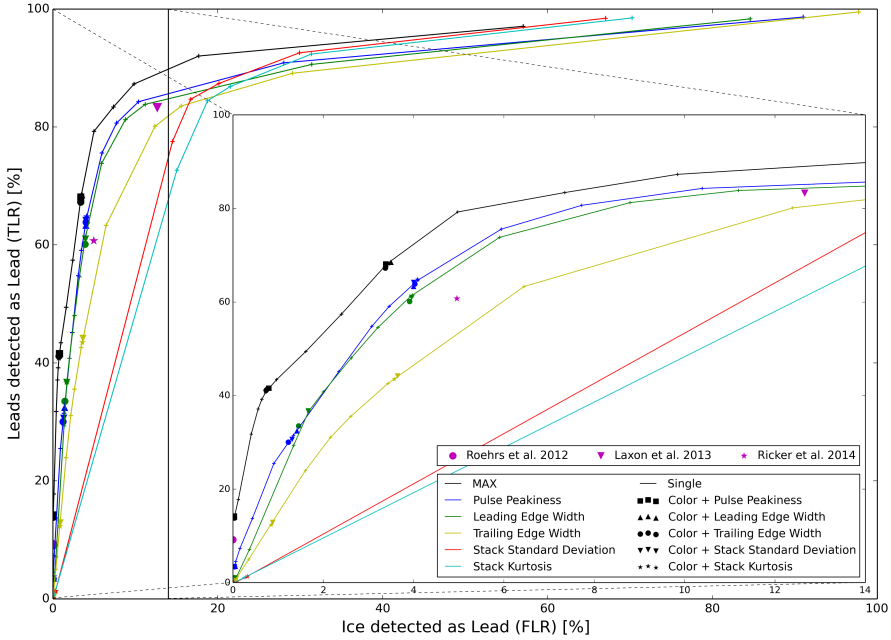


Figure 4. ROC graph of tested classifiers with altering thresholds (Θ) on one (connected by lines) and two (marker) parameters as well as predefined classifiers (magenta markers). In the two dimensional case the color indicates one of the parameters and the shape the other one. The insertion is a zoom on small False Lead Rates.

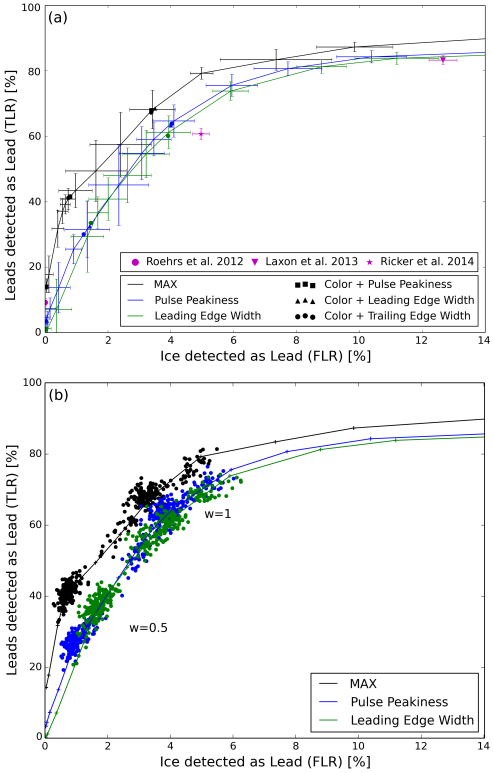


Figure 5. (a) ROC graph including error estimates in terms of SD of the 200 runs for each weight of the single feature classifiers using the MAX, PP and LEW as well as the performances and SDs of the predefined classifiers. For comparison the performances of selected two dimensional classifiers are included. **(b)** Performances of each individual run being part of the single feature classifiers using the MAX, PP and LEW with weights of 0.5 and 1 (dots) in combination with mean values for all weights (lines).

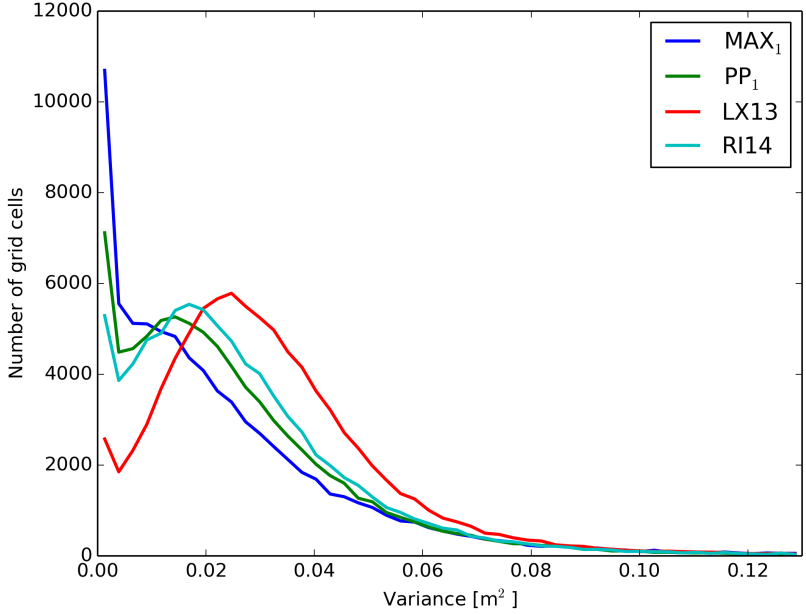


Figure 6. Histograms of grid cell SSH variance from different classifiers. Only values based on at least three lead detections are considered.

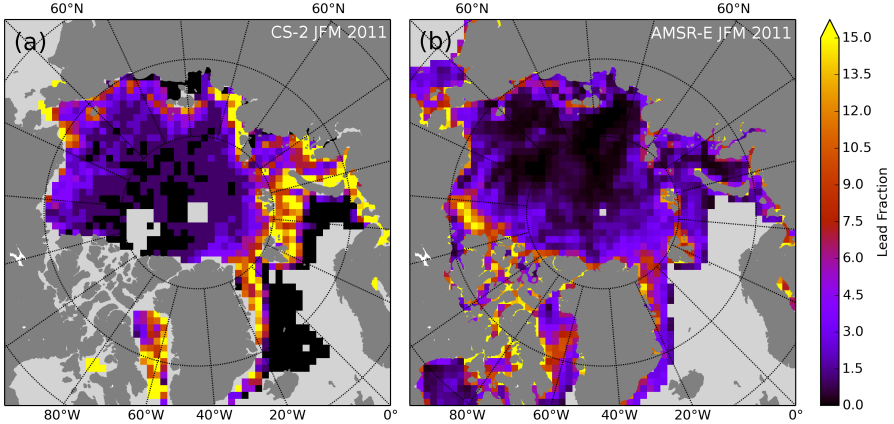


Figure 7. Lead fraction derived from CS-2 SAR-mode **(a)** and from Röhrs et al. (2012) **(b)** on a North Polar Stereographic Grid with a resolution of 99.5 km × 99.5 km, merged from January to March 2011. Only values based on at least 2000 CS-2 measurements north of 65° N **(a)** or with a grid cell data coverage of 10 % or more **(b)** are shown. Missing CS-2 estimates north of Canada are caused by the use of a different mode in the Wingham Box.

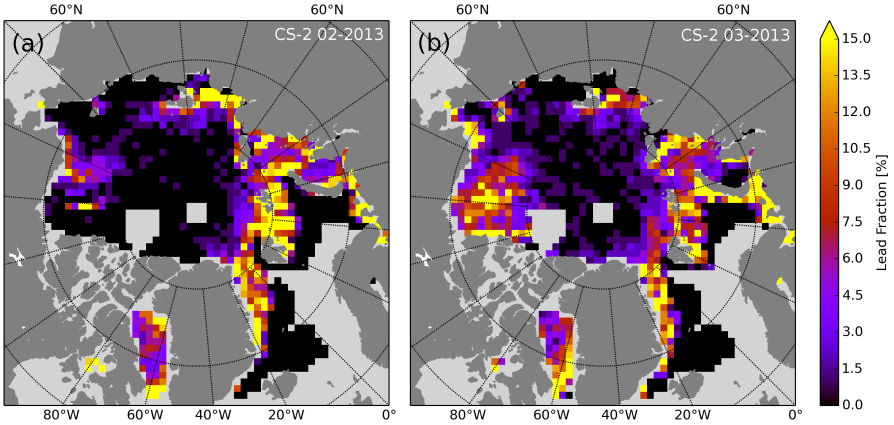


Figure 8. Lead fraction derived from CS-2 SAR-mode on a North Polar Stereographic Grid with a resolution of 99.5 km × 99.5 km from February **(a)** and March 2013 **(b)**. Only lead fraction values north of 65° N based on at least 1000 measurements are shown. Missing estimates north of Canada are caused by the use of a different mode in the Wingham Box.

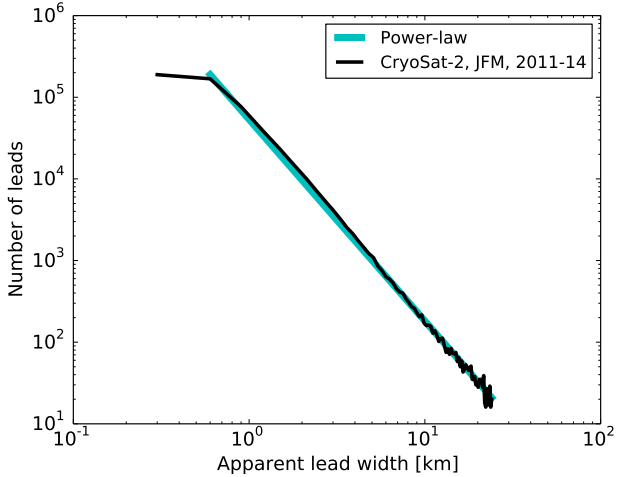


Figure 9. Apparent lead width distribution from all CS-2 SAR-mode ocean measurements north of 65° N in winter season (JFM) from 2011 to 2014. The distribution of a power-law with exponent of 2.47 is included for comparison, forming a straight line in a double logarithmic presentation. See text for definition of the apparent lead width.