<u>Reviewer #1</u>

Received and published: 10 August 2020

This manuscript uses RADARSAT-2 imagery to derive peak melt pond fraction values for sea ice in the Canadian Arctic Archipelago between 2009 and 2018. The basic method for deriving peak pond fraction was developed in an earlier publication, and this work applies that method to a larger dataset from a different satellite. The manuscript is well written and has only a few grammatical errors that are noted below. The results presented offer valuable insight into sea ice trends and variability in the CAA. However, there are a few issues with the validation of the RADARSAT-2 derived data that should be fixed or clarified prior to publication.

Howell et al.

We thank this reviewer for her/his comments that have improved this manuscript considerably. We have incorporated almost all of her/his suggestions.

Reviewer #1

General Comments

You define f_p as melt pond fraction. Throughout the paper you also use f_p to refer to peak melt pond fraction calculated from RADARSAT-2. It would improve clarity to separate the notation for these two different parameters.

Howell et al.

Very good suggestion. We have chosen to define peak melt pond fraction as f_{pk} and have changed the text throughout the manuscript to reflect this new notation.

Reviewer #1

There are two issues with the in-situ comparison:

1. The spatial footprint of the LIDAR scans from Landy et al., (2014) are small in comparison to the 100m resolution of RADARSAT-2 data used. These in-situ datasets would only cover 1-2 pixels in the radar image. Does this area represent the whole region? Perovich (2002) determined the aggregate scale (area at which a sample can be considered representative of the larger region) at SHEBA to be multiple kilometers. If the aggregate scale is much lower in the CAA (more homogeneous ice cover) this should be discussed.

Howell et al.

It is true the LiDAR areas would cover only ~1-2 pixels, however we only compared the LiDAR pond fraction to the ~1-2 RADARSAT-2 pixels directly coincident with the site. Therefore, we are not validating RADARSAT-2 melt pond fraction against a spot LiDAR *in situ* measurements, we are just validating the entire 100 m LiDAR melt pond fraction directly at the sampling site. In this case, it does not matter whether the *in situ* samples are representative of the aggregate scale. We have clarified this in text so other readers to confuse other readers:

Revised Section 3.2

Figure 7a compares the time series of the entire 100 m LiDAR f_k coincident with the f_{pk} determined from RADARSAT-2 at the coinciding pixels.

Reviewer #1
2. Two in-situ samples are not enough to assess the accuracy of this method given the error presented in Figure 7. Here the prediction for 2011 is correct and the prediction for 2012 is not. On line 180 you state that the error is 0.1, but it looks more like 0.2 in the figure. Have you considered other in-situ datasets? For example, the three years of melt pond fraction timeseries observed on landfast ice near Utqiagvik, AK described in Polashenski et al., (2012)?

Howell et al.
We are limited by the scarcity of *in situ* melt pond fraction observations in the CAA and would have used more if we could. Moreover, finding observations that coincide with peak pond fraction further adds to the scarcity problem and the MODIS analysis was attempt to alleviate this problem. Unfortunately, we cannot use the *in situ* melt pond fraction dataset from Polashenski et al. (2012) because our RADARSAT-2 data only has consistent coverage in the Canadian Arctic waters in accordance with the operational domain of the Canadian Ice Service and therefore the Chukchi Sea is not covered. Despite having only two *in situ* samples, they least cover a long temporal time period allowing us to test whether RADARSAT-2 picks out the seasonal mean pond fraction or peak pond fraction. However, we do have aerial photograph estimates of melt pond fraction obtained over and adjacent the LiDAR site in 2012 from Scharien et al. (2014), which we have made use of to compare with RADARSAT-2 f_{pk} estimates. We have added a new Figure 7b with the aerial photograph data and revised the following sections:

Revised Section 3.2

Figure 7a compares the time series of the entire 100 m LiDAR melt pond fraction coincident with the f_{pk} determined from RADARSAT-2 at the coinciding pixels. For 2011, RADARSAT-2 f_{pk} corresponds to the end of stage I and beginning of stage II thus providing a very good representation of the seasonal peak of the f_p, when the melt pond control on heat uptake and ice decay, through the ice-albedo feedback, is greatest. For 2012, RADARSAT-2 f_{pk} also corresponds to the end of stage I and beginning of stage II but is ~0.1 lower than *in situ* f_p values. This is likely due to the short duration but very high maximum f_p of 0.78 in 2012 as Scharien et al. (2017) found that equation (1) sometimes underestimates very high f_p due to the low γ^{o} signal associated with very smooth FYI.

Figure 7b shows the distribution of RADARSAT-2 f_{pk} and the f_p determined from aerial photo observations on June 22nd, 2012 near Resolute. The aerial photographs were acquired within 1 week of f_{pk} coverage being observed at the LiDAR site. The comparison was done by averaging all RADARSAT-2 pixels within each aerial photo. The mean aerial photograph f_p was 0.54 and RADARSAT-2 f_{pk} was 0.53 with an the RMSE of 0.10 and bias of 0. The distributions are in reasonably good agreement but RADARSAT-2 values are slightly narrower than the distribution of f_p from the aerial photographs. It is likely the RADARSAT-2 distribution is narrow on the left tail because our method captures peak pond coverage and some of the regions photographed were before or after their seasonal peak. We attribute the narrow right tail to the documented underestimation of equation (1) from Scharien et al. (2017). However, it is notable that both RADARSAT-2 and the aerial photograph datasets capture the same bimodal f_p distribution, with the first mode around 0.4-0.5 characterizing rougher sea ice areas and the second mode around 0.7 capturing smooth flooded sea ice.

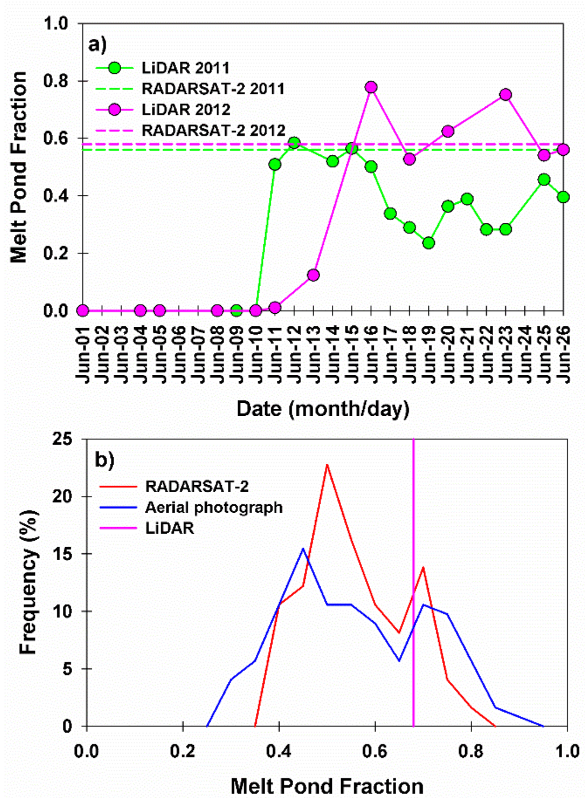

Figure 7. a) Temporal evolution of observed melt pond fraction (f_p) and RADARSAT-2 peak melt pond fraction (f_{pk}) at *in situ* observations sites for 2011 (74.7229°N; 95.1763°W) and 2012 (74.7264°N; 95.5772°W). b) Frequency distribution of RADARSAT-2 f_{pk} and aerial photograph f_p observations in Resolute Passage on June 22, 2012; the pink vertical link represents the mean LiDAR f_p on June 22, 2012.

Revised Section 2.1

Aerial photographs of estimated f_p directly over the LiDAR site and the adjacent sea ice area away from land and open water were also obtained on June 22, 2012. The aerial photographs have a pixel resolution 0.22 m resolution, cover 750 m by 750 m. In total, 123 aerial photographs of f_p were used and a complete description of the dataset is provided in Scharien et al. (2014).

Added Reference

Scharien, R. K., Hochheim, K., Landy, J., and Barber, D. G.: First-year sea ice melt pond fraction estimation from dual-polarisation C-band SAR – Part 2: Scaling in situ to Radarsat-2, The Cryosphere, 8, 2163–2176, https://doi.org/10.5194/tc-8-2163-2014, 2014.

Reviewer #1

Lines 183-194: What is the conclusion from the comparisons with MODIS? You note the reasons why RADARSAT-2 derived f_p and MODIS f_p could be misaligned (i.e. that the MODIS product is an 8-day average and peak ponding occurs on short timescales), and I am left with the impression that the MODIS data do not agree with your results. I would suggest expanding or clarifying the statistical analysis here. In Figure 8, both 2010 and 2011 make the RADARSAT-2 look statistically different than MODIS. The mean (blue line) of RADARSAT-2 is approximately equal to the max (top whisker) of MODIS.

Howell et al.
This is a good point raised by the Reviewer and we were not definitive in our wording based on the boxplots. The conclusion is that RADARSAT-2 pond fraction is higher on average than MODIS because the MODIS 8-day product is not representative of f_{pk} in the CAA. The weekly boxplots and max MODIS pond fraction boxplot all support this conclusion. We note that the box plot of maximum f_p from MODIS does capture some regions at peak during the 8-day time series. Another point is that MODIS estimation error needs to be acknowledged because although it is treated here as validation for the RADARSAT-2 f_{pk} estimates and rightly so but it has its own error component. We clarify this section in here as follows:

Revised Section 3.2
The seasonal time series of the 8-day composite MODIS f_p, the maximum seasonal MODIS f_p and the predicted RADARSAT-2 f_{pk} for 2009-2011 is shown in Figure 8. MODIS f_p observations within the CAA indicate initial pond formation occurred in May for all years with f_{pk} reached in mid-July for 2009 and in early June for 2010 and 2011. Compared to the RADARSAT-2 f_{pk} values, the peak MODIS f_p is ~0.20 smaller. RADARSAT-2 f_{pk} is higher on average than MODIS because the MODIS 8-day product does not represent f_{pk}. The MODIS f_p observations are determined weekly using 8-day composite image products that would include some melt pond formation and drainage processes prior-to, and after, the seasonal peak. Moreover, MODIS f_p observations give the time series of f_p therefore even the highest seasonal estimated MODIS f_p is reduced because while some regions of the CAA are at their seasonal peak but others are behind or ahead. To that end, we also calculated the maximum f_p from MODIS regardless of timing during the melt season, for each pixel, also shown in Figure 7. These values more closely compare with the RADARSAT-2 f_{pk} but are still ~0.05 smaller on average. Even the maximum f_p from MODIS is from an 8-day running mean of daily pond fraction estimates, so will underestimate the f_{pk} if the duration of peak ponding is <8 days. However, the top whisker of the box plot of the maximum f_p from MODIS indicates that MODIS does capture some regions at peak during the 8-day time series. Although we are using MODIS f_p product to compare against our RADARSAT-2 f_{pk} estimates, Rösel et al. (2012) found that the MODIS f_p product also has errors up to ~0.1. Overall, MODIS f_p estimates are more representative of the seasonal mean f_p rather than f_{pk} within the CAA.

Revised Conclusion
Based on our comparative analysis, RADARSAT-2 f_{pk} is more representative of peak f_p within the CAA compared to the MODIS 8-day product which on average was found to underestimate f_{pk} by ~0.2 and the is more representative of the seasonal mean f_p.

Reviewer #1
Specific comments
104 – Maybe this is covered in the Scharien paper, but is there a hypothesis for why this correlation exists? Is this method essentially just relating surface roughness (via radar backscatter) to peak pond fraction?

Howell et al.
Yes, it is explicitly covered and exploits the basic hypothesis that winter backscatter increases with increasing topography, for FYI, and increasing volume scattering, which is related to topography, for MYI. In each case, the increased topography leads to lower pond fraction, and

visa versa. The high resolution optical imagery helps exploit this relationship. That is, using high spatial resolution optical imagery Scharien et al. (2017) were able to isolate internally coherent, and externally discrete, zones of sea ice in order to compare backscatter/texture and f_p and thus create simple models.

Reviewer #1
107 – If fp is calculated directly from each radar pixel value (Eqn. 1), how does speckle filtering impact the fp results?

Howell et al.
The impact of speckle filtering/not filtering was not assessed. As with most SAR images speckle is a problem with the goal being to obtain the most representative backscatter value for a local region (i.e. a cleaner image). The Lee facilitates this by smoothing the image without removing edges or sharp features in the images while minimizing the loss of radiometric and textural information. Although speckle filtering will change the f_{pk} results for specific pixels, it will not impact f_{pk} at the scale of the filter (i.e. within an x by x pixel area).

Reviewer #1
165 – If both sensors are the same frequency, why is there any difference here (Figure 6) (spatial resolution difference? Sensor measurement errors?)

Howell et al.
Good point. We should have provided some explanation for these differences

Revised Section 3.2

Frequency distributions of RADARSAT-2 f_{peak} and Sentinel-1 f_{peak} from Scharien et al. (2017) in the CAA for 2016 and 2017 are shown in Figure 6. Sentinel-1 appears to estimate more regions of lower f_{peak} compared to RADARSAT-2 which are typically associated with MYI. Whereas, RADARSAT-2 estimates more regions of higher f_{peak} which are typically associated with smooth FYI. We consider these subtle differences to be primarily the result of taking the mean of all available April RADARSAT-2 imagery (Table 1) over all incidence angles in the CAA compared to only using images from Sentinel-1 within the CAA constrained to a certain incident angle range. As shown in Figure 2, the uncertainty in RADARSAT-2 f_{pk} varies depending on the number of pixel overlaps (images). Overall, the f_{pk} distributions are in good agreement between both sensors.

Reviewer #1
180 – this looks like it is 0.2 lower (difference between dashed pink line and peak pink dot). Am I reading this plot incorrectly?

Howell et al.
It should be 0.19 not 0.9. We have revised it ~0.2.

Reviewer #1
248 – “Slightly lower” is maybe an understatement? It is 20% lower. Either way, quantify the amount it is lower here.

Howell et al.
Revised Section 3.2
RADARSAT-2 f_{pk} was found to be in good agreement with the f_p maximum extent observed *in situ* for 2011 but was ~0.2 lower than 2012 when f_{pk} was very large (> 0.7) for a very short duration (1-2 days).

Reviewer #1
251 – In 214-231 you posit that the predictive power of this method only holds for landfast ice (i.e. when ice breakup is due to thermodynamics and not due to ice motion), how would this method be applicable to pan-Arctic estimates?

Howell et al.
In that case a Lagrangian tracking approach would be needed or the integrated melt pond fraction could be used with evolving sea ice extent. In both cases, significant testing would be required. We are working on this, but it is considerably outside the scope of this analysis.

Reviewer #1
Technical Corrections
59-61 – Run-on sentence.

Howell et al.
Revised Introduction
Model simulations have been utilized to understand the current and predicted future variability of sea ice conditions in the CAA (e.g. Dumas et al., 2006; Sou and Flato, 2009, Howell et al., 2016; Laliberté et al., 2016; Hu et al., 2018; Laliberté et al., 2018). However, modeling the CAA still remains challenging because complex sea ice dynamic and thermodynamic processes are often not accurately resolved in its narrow channels and inlets

Reviewer #1
97 – "during April in within the CAA": Extra “in” here.

Howell et al.
Removed

Reviewer #1
152 – This sentence is unclear.

Howell et al.
Revised:
What is interesting in Figure 5a is that the mean RADARSAT-2 f_{peak} in 2009 was lower than all years from 2014-2018 (with the exception of 2016) despite the CAA containing less MYI area.

Reviewer #1
154 – “in addition” and "also" are redundant here.

Howell et al.
Removed “also”

Reviewer #1
161 – 3.2 header has extra "and". Also consider including oxford comma in this list for added clarity.

Howell et al.
Revised:
3.2 Comparison of RADARSAT-2 f_{pk} with Sentinel-1 f_{pk}, in situ f_p, and MODIS f_p

Reviewer #1
183 – Again a stylistic choice, but I find oxford commas to be helpful for clarity.

Howell et al.
Revised:
The seasonal time series of the 8-day composite MODIS f_p, the maximum seasonal MODIS f_p, and the predicted RADARSAT-2 f_{pk} for 2009-2011 is shown in Figure 8.

Reviewer #1
190 – "but" is an extra word here.

Howell et al.
Removed.

Reviewer #1
192 – Do you mean Figure 8 here?

Howell et al.
Yes. Changed to Figure 8.

Reviewer #1
215 – "The origin of the some of the ice” extra words here.

Howell et al.
Yes. Removed “the some of”.

Reviewer #1
239 – "Overall, within the: : : ": Revisit sentence structure here.

Howell et al.
Revised:
Overall, within the Viscount-Melville Sound region of CAA there is a period for which a significant statistical relationship exists between RADARSAT-2 f_{pk} and the summer ice area before sea ice dynamics degrades the relationship.

Reviewer #1
253 – "Was found to be excellent agreement": Missing "in" here.

Howell et al.
Added “in”.

Reviewer #1
249 – "maybe” should be "may be" in this context.

Howell et al.
Changed.