



Jie Su (Orcid ID: 0000-0002-4342-1185)

Landy Jack, Christopher (Orcid ID: 0000-0002-7372-1007)

A new algorithm for sea ice melt pond fraction estimation from high-resolution optical satellite imagery

Wang Mingfeng^{1,2,3}, Su Jie^{1,2,3*}, Jack Landy⁴, Matti Leppäranta⁵, Guan Lei^{2,6}

¹Physical Oceanography Laboratory/CIMST, Ocean University of China, Qingdao, Shandong, China

²Qingdao National Laboratory for Marine Science and Technology, Qingdao, Shandong, China

³University Corporation for Polar Research, Beijing, China

⁴Bristol Glaciology Centre, University of Bristol, Bristol, UK

⁵Department of Physics, University of Helsinki, Helsinki, Finland

⁶Department of Marine Technology, College of Information Science and Engineering/Institute for Advanced Ocean Study, Ocean University of China, Qingdao, Shandong, China

Key points:

- Sentinel-2 and IceBridge imagery are used to show that algorithms assuming fixed melt pond-reflectance largely underestimate melt pond fraction.
- A new algorithm based on the polar coordinate transformation that treats melt ponds as variable-reflectance has been developed.
- The accuracy and precision of the new Sentinel-2 algorithm, assessed with IceBridge DMS imagery, are higher than previous algorithms.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019JC015716

Abstract

Melt ponds occupy a large fraction of the Arctic sea ice surface during spring and summer. The fraction and distribution of melt ponds have considerable impacts on Arctic climate and ecosystem by reducing the albedo. There is an urgency to obtain improved accuracy and a wider coverage of melt pond fraction (MPF) data for studying these processes. MPF information has generally been acquired from optical imagery. Conventional MPF algorithms based on high-resolution optical sensors have treated melt ponds as features with constant reflectance; however, the spectral reflectance of ponds can vary greatly, even at a local scale. Here we use Sentinel-2 imagery to demonstrate those previous algorithms assuming fixed melt pond-reflectance greatly underestimate MPF. We propose a new algorithm (“LinearPolar”) based on the polar coordinate transformation that treats melt ponds as variable-reflectance features and calculates MPF across the vector between melt pond and bare ice axes. The angular coordinate θ of the polar coordinate system, which is only associated with pond fraction rather than reflectance, is used to determine MPF. By comparing the new algorithm and previous methods with IceBridge optical imagery data, across a variety of Sentinel-2 images with melt ponds at various stages of development, we show that the RMSE value of the LinearPolar algorithm is about 30% lower than for the previous algorithms. Moreover, based on a sensitivity test, the new algorithm is also less sensitive to the subjective threshold for melt pond reflectance than previous algorithms.

Plain Language Summary

Melt ponds are pools of open water that form on the sea ice surface in the Arctic in summer months. Sea ice covered by melt ponds absorbs more solar heat than solid ice which speeds up the rate of sea ice melt. To study this process, we need melt pond coverage data that is generally obtained from satellite observations. There are several methods to calculate the melt pond fraction from high-resolution optical satellite imagery but these methods assume that melt ponds have constant optical properties, which is not true. In this study, we show that these existing methods significantly underestimate melt pond fraction. To address this problem, we present a new method that treats melt ponds as features having variable optical properties. Images of sea ice from aircraft are used to assess the new algorithm and two previous methods, demonstrating that the new algorithm has higher accuracy and precision.

Key words: Melt ponds, Sea ice, Remote sensing, Arctic

1. Introduction

Melt ponds form on the sea ice surface during the summer. The albedo of melt ponds is greater than that of open water but significantly smaller than that of snow and ice. As a result, melt ponds play an important role in the positive albedo feedback mechanism of sea ice, where a decrease in the area of sea ice lowers the albedo that the loss of sea ice area (Curry, 1995). Furthermore, refrozen melt ponds conduct sensible heat and release latent

heat in winter, slowing down sea ice growth and regulating the ice mass balance (Taylor et al., 2004). By reducing the albedo, melt ponds allow solar radiation to be transmitted more easily through sea ice, thus heating the oceanic mixed layer and forming a more productive environment for phytoplankton (Palmer et al., 2014). Melt pond coverage varies rapidly and widely throughout the melt season in a given place and can vary substantially from one place, or ice type, to another. Observations have shown that the pond fraction can vary by more than 60% over the course of a melt season and by up to 40% between years at the same location (Polashenski et al., 2012; Landy et al., 2014; 2015). These facts emphasize the urgency of obtaining reliable MPF data.

The basic methods for retrieving MPF from optical satellite imagery are unmixing the pond coverage from mixed pixels, or separating melt pond pixels from open ocean (i.e., leads) and snow/bare-ice pixels. The latter method is dependent on the resolution of the data; to be detected, the diameter of a melt pond must be larger than the pixel size. The size distributions of Arctic melt ponds are dominated by small ponds ($<200\text{m}^2$) (Perovich et al., 2002; Tschudi et al., 2001; Lu et al., 2010). The observations from the Surface Heat Budget of the Arctic Ocean (SHEBA) program (between 20 May and 4 October 1998) illustrate that the size of melt ponds is in the range $1\text{-}300\text{ m}^2$. Observations during CHINARE-2008 (between 17 Aug and 5 September 2008) captured melt ponds with area up to 10^3 m^2 . A combination of high- and medium-resolution sensors can be used to acquire a melt pond dataset of the entire Arctic (Rösel and Kalesche, 2011). This would represent a trade-off between the wide coverage and daily repeat frequency of medium-resolution sensors such as MODIS (500 m resolution), which require pixel unmixing, and the narrow coverage and weekly repeat-frequency of high-resolution sensors, such as Sentinel-2 (15 m resolution), which generally need unmixing but can resolve some large individual ponds with areas $>225\text{ m}^2$.

Several melt pond fraction retrieval algorithms have been developed for optical imagery, using MODIS (Tschudi et al., 2008; Rösel et al., 2012), MERIS (1000m resolution) (Zege et al., 2015) or Landsat 7 (30 m resolution), including the MPF determined from the triangle-based methods of Markus et al (2002) and principal component analysis (PCA)-based method of Rösel and Kalescheke (2011). The MODIS algorithms (Tschudi et al., 2008; Rösel et al., 2012) are based on a spectral unmixing procedure first proposed by Tschudi et al. (2008), where the reflectance of each surface type within each of the sensor's spectral bands is assumed to be constant. The MERIS algorithm (Zege et al., 2015) doesn't use fixed values for surface reflectance, but applies a physical model to derive them from the inherent optical properties of bare sea ice and melt ponds. A small number of these optical properties are free parameters and optimized in the MPF determination. The Landsat-7 algorithms are based on supervised classification (Rösel and Kalescheke, 2011). However, for these algorithms the melt ponds are assumed to have a fixed spectral signature, determined from training data, due to the simplicity of the method and limited resolution and quality of the Landsat-7 imagery. The assumption of a single fixed melt pond spectrum could present a significant limitation to the accuracy of available algorithms, because in situ studies have shown that the optical properties of melt ponds vary

considerably (Istomina et al., 2016). The broadband albedo of ponds can vary between 0.1 and 0.7 (and between 0.1 and 0.4 at red wavelengths), depending principally on the underlying ice thickness (Istomina et al., 2016; Lu et al., 2018a, b). There are, also, some MPF retrieval algorithms using Synthetic Aperture Radar (SAR) and passive microwave data, which have the advantage of all-weather and day plus night operation. However, currently operating SAR systems (Sentinel-1, RCM) do not have the polarimetric configurations required to fully exploit the developed MPF retrieval techniques, and cannot yet be reliably applied over multi-year ice (Yackel et al., 2000; Makynen et al., 2014; Scharien et al., 2014a, b; Li et al., 2017; Tanaka et al., 2016).

In this study, we first illustrate the errors that can be introduced by available melt pond fraction algorithms owing to their limited premise of fixed melt pond spectral reflectance signatures. We then present a new computer-vision based technique that better accounts for the melt pond spectral reflectance variation and identifies dark ponds, and propose a new algorithm to correct this problem. A set of Sentinel-2 images were selected for the training and application of the new algorithm. Finally, aerial images from the NASA Operation IceBridge program are used to independently verify retrievals from the new algorithm.

2. Data

Optical imagery from the twin Sentinel-2 satellites and aerial images from the IceBridge airborne campaign are used in this study. The main research area is located to the west of the Canadian Arctic Archipelago, including Sentinel-2 scenes 1-5 (Figure 1). The acquisition times of the selected scenes are between mid-June and late July, and the latitudes range from 73°N to 83°N. Since the melting stage of Arctic sea ice varies with time and latitude, these selected scenes can ensure a representative cross-section of training and application data across different first-year ice (FYI) and multiyear ice (MYI) areas, ice thicknesses and stages of melt pond development (Landy et al., 2014). We selected 40 subsamples across the 5 scenes for MPF retrieval owing to their high melt pond fraction and absence of clouds. MPF determination results of scene 5 are verified with 35 coincident aerial images from the IceBridge flight on 24th July 2017 (Figure 1).

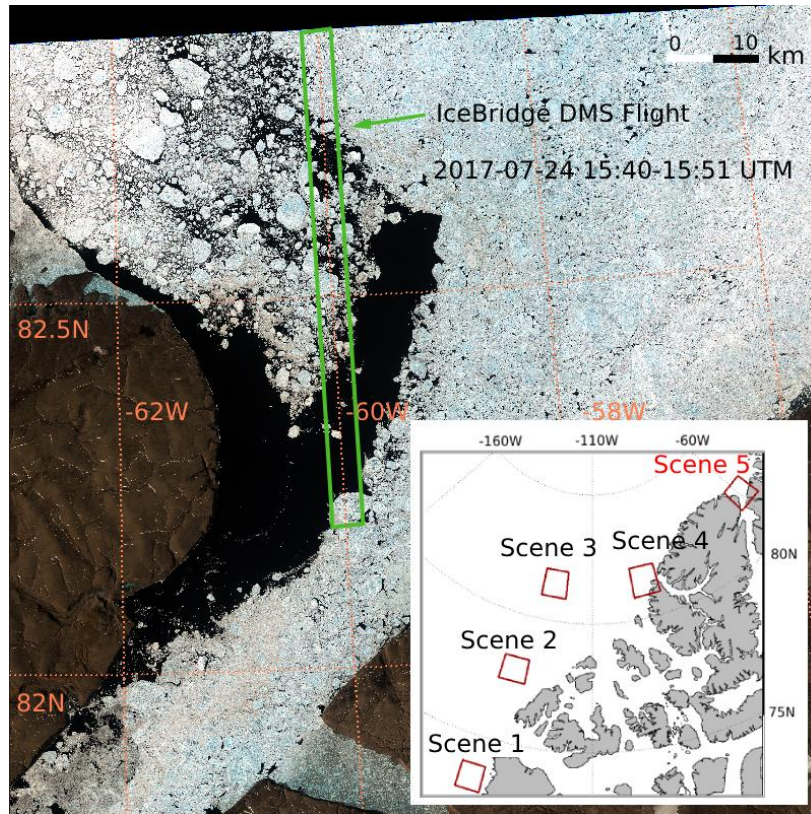


Figure 1. Overview of the study area with the locations of the selected Sentinel-2 scenes (scene 1-5 with red frames). Scene 5 is used for validation with IceBridge aerial images (green frame). The acquired times of Scene 1-5 are 29 July 2016, 4 July 2017, 29 June 2017, 12 June 2017, and 24 July 2017, respectively.

The Sentinel-2 satellites are part of the Copernicus program of the European Space Agency (ESA). They carry wide-swath high-resolution multispectral imagers with 13 spectral bands in the visible, near-infrared, and short wave infrared part of the spectrum. The visible and near-infrared bands used in this study have a pixel size of 10m. The mission is based on a constellation of two identical satellites in the same orbit, 180° apart for optimal coverage and data delivery (Bindschadler et al., 2008; Drush et al., 2012).

We use the IceBridge Digital Mapping System (DMS) Level-1B Geolocated and Orthorectified images (Dominguez, 2010) in this study as an independent high-resolution reference to Sentinel-2 melt pond discrimination. The main components of this system include a 21-megapixel digital camera, a computer-controlled intervalometer, and a POS/AV precision orientation in-flight operator. The DMS images are available as true color imagery (3 bands in RGB), with a nominal spatial resolution of 0.1 m, but note that the 3-band images cannot be analyzed with an identical method to the 13-band Sentinel-2 imagery.

The MPF determined from the triangle-based methods of Markus et al. (2002) and PCA (Principal Component Analysis)-based method of Rösel and Kalescheke (2011) were originally designed for Landsat 7 data. Although the band intervals of Sentinel-2 are similar

to Landsat 7, they are slightly different in the wavelength response range. In these two algorithms, we used band 1 (450-520 nm), band 2 (520-600 nm), band 3 (630-690 nm) and band 4 (770-900 nm) of Landsat 7, while the corresponding bands in Sentinel-2 are band 2 (440-538 nm), band 3 (537-582 nm), band 4 (646-684 nm) and band 8 (760-908 nm). Such differences will not have a significant influence on the application of Landsat 7-based algorithms to Sentinel-2.

3. Methodology

3.1 Selection of training areas

We selected ten training areas for each surface class (melt pond, ice, lead) within each of the 5 scenes shown in Figure 1. These training areas contained a large spread of spectral properties including, for example, melt ponds with visible colors from bright to dark blue. The mean and standard deviation of the spectral reflectances of these training areas are shown in Table 1. We use the reflectances from these training areas to parameterize all of the algorithms deployed here for estimating MPF.

Table 1. Spectral reflectance of the surface type used in the new algorithm

Sentinel-2 Band	Bandwidth (nm)	Melt pond (mean +/- std dev)	Bare ice (mean +/- std dev)	Snow cover (mean +/- std dev)	Lead (mean +/- std dev)
Band 2	440-538	0.46±0.12	0.64±0.072	0.77±0.057	0.1±0.044
Band 3	537-582	0.35±0.13	0.57±0.073	0.69±0.059	0.08±0.041
Band 4	694-713	0.23±0.16	0.57±0.083	0.66±0.072	0.05±0.041
Band 8	760-908	0.13±0.13	0.49±0.086	0.51±0.077	0.026±0.032

We selected one of these training areas to illustrate the limitations of former MPF retrieval algorithms and the improvement of the new algorithm (Figure 2a). This area is chosen because it includes two melt ponds with distinctly different color saturation (melt pond 1: bright blue, and melt pond 2: dark blue), which is principally caused by different ice thicknesses under the ponds (Istomina et al., 2016; Peng Lu et al., 2017). Color saturation (here refer to the intensity of blue color) in Figure 2d is therefore used to quantify the differences between these two types of pond, giving the intensity of reflectance in the blue channel. Some large individual ponds can be found in Sentinel-2 imagery while most of ponds exist in mixed pixels. Snow-cover and bare ice classes were selected to represent the spectral variation of different types of pure ice. Snow-cover areas are taken from the highest reflectance pure ice pixels, and bare ice areas are taken from the lowest reflectance pure ice pixels. As there are no reference data to verify the surface type, the naming here only refers to two variants of sea ice cover with distinctly different optical properties.

In order to test the unmixing ability of different algorithms, two classes of mixed pixels were also selected as training areas: mixed pixel 1 is composed of melt pond 1 and snow-covered ice, and mixed pixel 2 is composed of melt pond 2 and snow-covered ice. The mixed pixels are located along the edges of melt pond 1 and 2. We made the best possible

estimate for pure ice and pure melt pond training pixels based on their spectral behavior (from the 13 bands of sentinel-2).

The MPF determined from the triangle-based methods of Markus et al. (2002) and the PCA-based method of Rösel and Kalescheke (2011) are given in Figures 2b, 2d. The two algorithms correctly identify melt pond type 1 (bright blue), whereas they failed to identify melt pond type 2 (dark blue); pond type 2 has only have calculated an MPF value at around 50% on average. In the following sections, we will show that this issue is pervasive and explains why these two algorithms frequently underestimate the true melt pond coverage on sea ice.

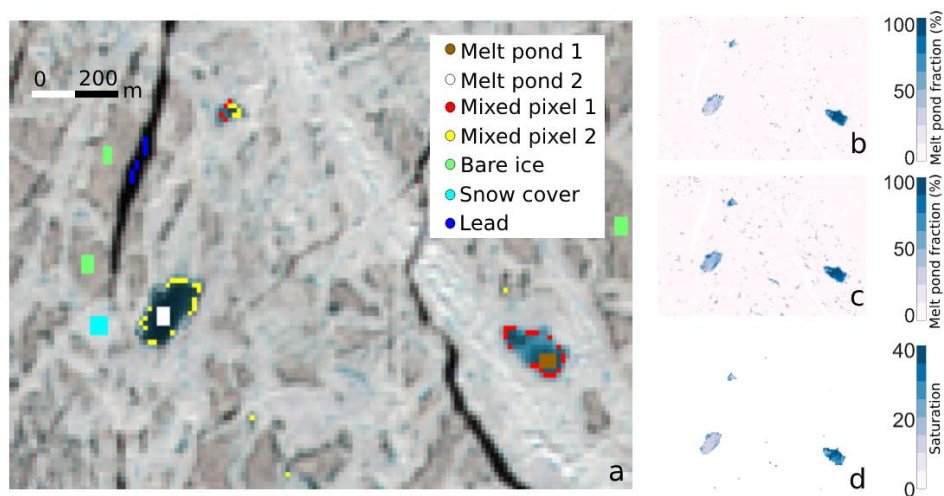


Figure 2. (a) True-color (band combination 4-3-2) subset from the Sentinel-2 scene displayed in Figure 1 with selected training areas. (b) MPF result using the Markus et al. (2002) algorithm. (c) MPF result using the PCA method. (d) The color saturation of melt ponds. Leads (open water) are marked with white color in (b), (d).

3.2 Tests of former algorithms

3.2.1 PCA algorithm

The PCA algorithm (Rösel and Kalescheke, 2011) is based on band 1 and band 4 of Landsat 7. In the first step, a two-dimensional (2-D) scatterplot of these two bands is displayed. The scatterplot includes two clusters: one composing pixels of open water and snow-covered ice, another composing pixel with a melt pond fraction. To separate the pixels with melt pond from other pixels, the second step, PCA transformation is applied on the 2-D scatterplot. The two uncorrelated components (named G2 and G1) are acquired from the PCA, and melt pond pixels are scaled between 0 and 100% MPF as a function of the G1 value (Rösel and Kalescheke, 2011).

Here we have applied the PCA algorithm to the subset of the Sentinel-2 scene in Figure 2a. The two-dimensional scatterplots (Figure 3a) of the subset including training areas illustrate

the first step of the PCA algorithm, in which band 2 (corresponding to band 1 of Landsat 7) and band 8 (corresponding to band 4 of Landsat 7) of the Sentinel-2 sensor are used. Equivalent to the scatterplots presented by Rösel and Kalescheke (2011) using Landsat 7, the main one-to-one axis includes classes of bare ice, snow-covered ice and leads (open water). Off the main axis, the remaining pixels include another axis of melt pond pixel having near-zero band 8 reflectance but band 2 reflectance from 0.3 to 0.5. It is noteworthy that the two melt pond training classes clusters at different points on the horizontal axis.

The two-dimensional scatterplot of the two uncorrelated components G2 and G1 (Figure 3b) shows that, after PCA transformation, the pixels with a melt pond fraction (from 0 to 100%) can be discriminated from water or bare-ice pixels by thresholding G1. Here we assign 100% pond fraction as the point where G1 equals 0.33. But similar to Figure 3a, the two melt pond clusters differ from each other distinctly in the horizontal axis (G1). Since melt pond fraction is scaled by the G1 value in the PCA algorithm, it correctly calculate an MPF value of around 100% to melt pond 1, but greatly underestimates the MPF of melt pond 2, as shown in Figure 2c. Although the choice of the '100% MPF point' is subjective, wherever it is chosen leads to either over- or under-estimation of a selection of other ponds. Based on the distribution of the two mixed pixel clusters, mixed pixels 1 and mixed pixels 2 have calculated an MPF of 65% and 38%. Although the melt ponds 1 cluster has been manually determined as having, in reality, 100% MPF, it has only calculated an MPF value of 27% by the PCA algorithm, less than the two classes of mixed pixels. This suggests an underestimation of the MPF in pixel of melt pond 1 by 83 pp (percentage points). As we have no independent verification data for this Sentinel-2 scene, here we can only assess the algorithm based on the relative MPF difference of different surface classes.

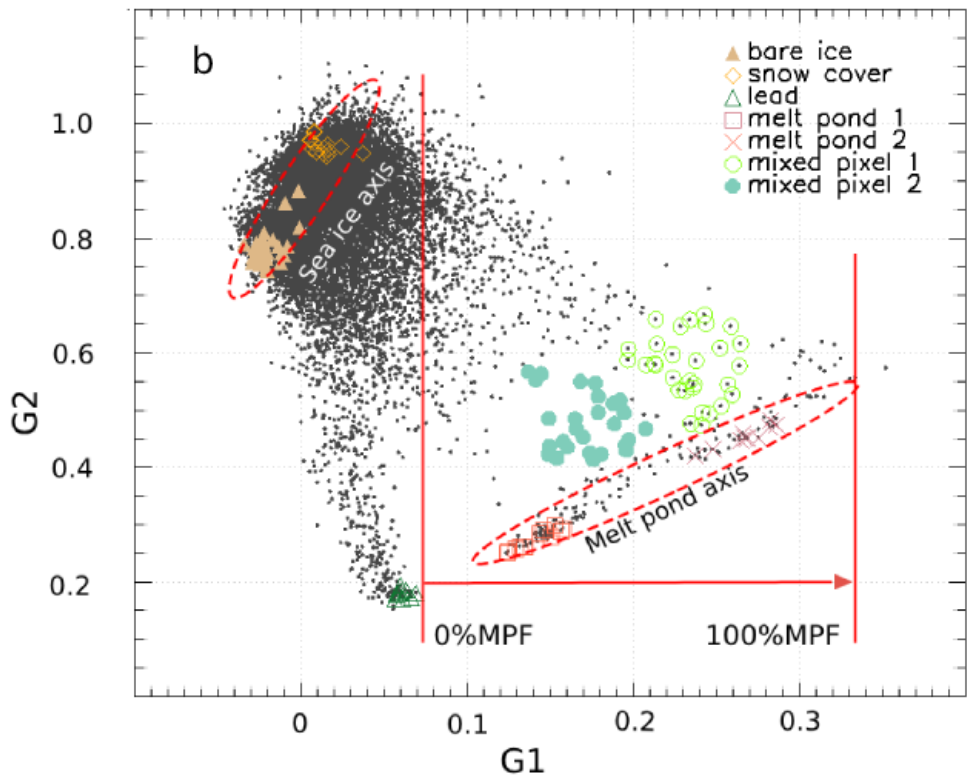
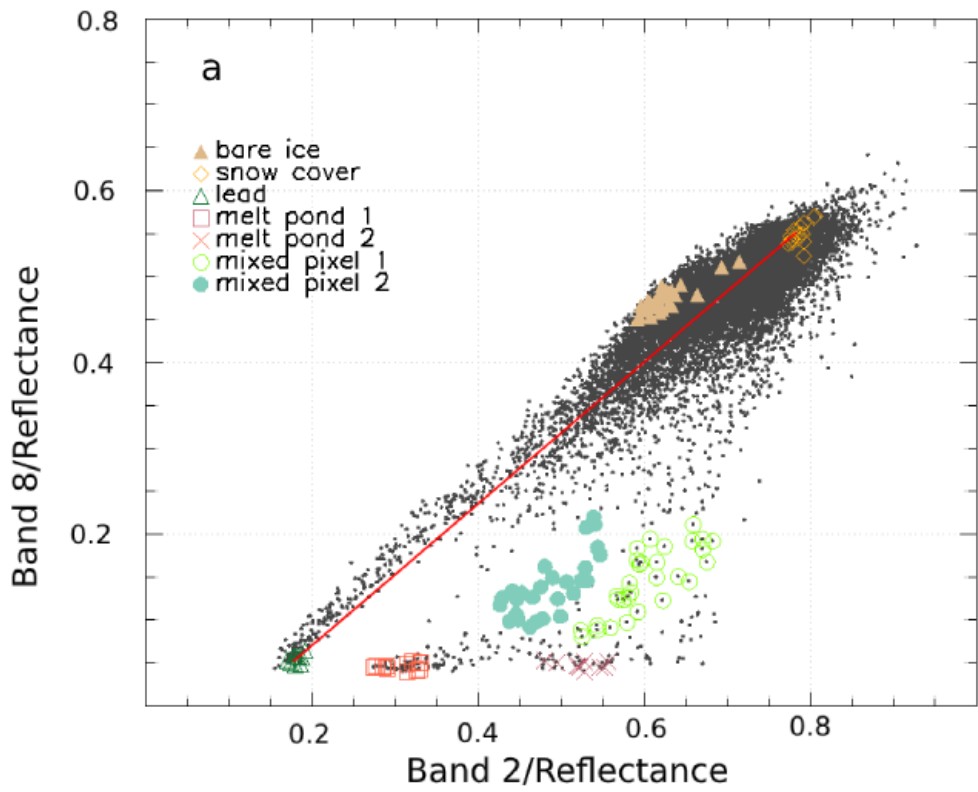


Figure 3. (a) 2-D scatterplots of Sentinel-2 reflectances of band 2 vs band 8 for the selected subset in Figure 2a. (b) Transformed new coordinate system with axes of G2 and G1. The scatter distribution of selected areas (In figure 2a) is illustrated with different markers and black dots refer to the coordinate value of unselected pixels in figure 2a)

3.2.2 Markus algorithm

For the MPF retrieval method of Markus et al. (2002), band 1 and the difference between band 2 and band 3 of Landsat 7 are used to detect melt ponds. The algorithm regards the three typical classes (snow/ice, open water, melt pond) to be distributed across a triangle between these band combinations. Pixels with a melt pond fraction from 0 to 100% are distributed along the vector connecting snow/ice and melt pond clusters.

Figure 4 illustrates the distribution relationship of the three typical classes using band 2 (corresponding to band 1 of Landsat 7) and the difference of band 3 (corresponding to band 2 of Landsat 7) and band 4 (corresponding to band 3 of Landsat 7) of the Sentinel-2 sensor. The open water and snow/ice nodes of the triangle are the same as Markus described, but the melt pond node is not well defined, with the melt pond 1 cluster located between the lead and dark pond clusters.

Like the PCA algorithm, the coordinates of the melt pond 1 and 2 clusters are significantly different from each other. As illustrated in Figure 4, if melt pond 1 is taken as the melt pond node of the triangle, melt pond 1 will be correctly determined, but the pond fraction of melt pond 2 will be underestimated, as shown in Figure 2b. If melt pond 2 is taken as the melt pond node of the triangle, the pond fraction of melt pond 1 is underestimated and some of the mixed pixels would have overestimated MPF.

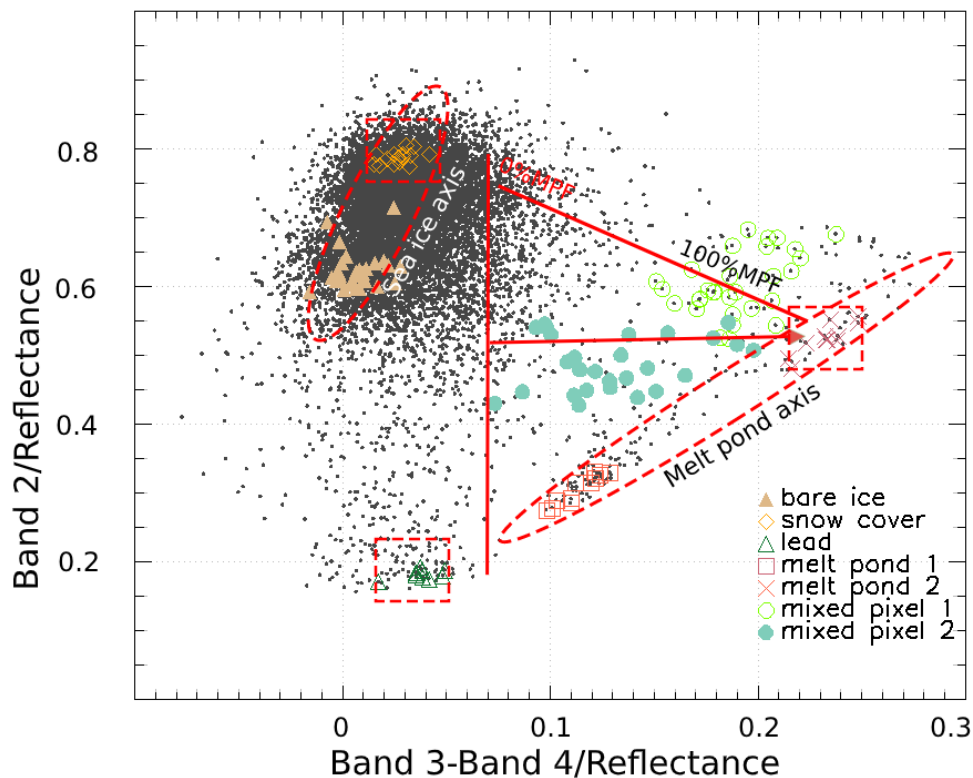


Figure 4. 2-D scatterplots of Sentinel-2 band 3 – band 4 reflectance difference vs band 2 reflectance for the selected subset scene in Figure 2a. The red boxes indicate the nodes of the triangle. Black dots refer to the coordinate value of un-selected pixels in Figure 2a)

Furthermore, as Figure 3 and Figure 4 show, a selection of the mixed pixels has calculated a higher MPF than the selected training area of melt pond 2, the darker pond. For both algorithms, variations in melt pond reflectance lead to an underestimation of MPF in certain areas of the image. Therefore, we now introduce a new melt pond fraction detection algorithm which can account for variations in the reflectance of melt ponds without underestimating MPF.

3.3 New MPF technique: the LinearPolar algorithm

From Section 2 and Figures 3 and 4, we found that the class distributions of the two algorithms have several features in common: 100% MPF pixels lie along a straight line, and 100% ice pixels including the bare ice cluster and snow-covered ice cluster compose a main ‘axis’. Here, we name these two axes the ‘melt pond axis’ and the ‘sea ice axis’, respectively. Pixels with a melt pond fraction <100% lie between these two axes. Based on these observations, we apply a polar coordinate transformation to the distributions, with the center defined as the intersection of the melt pond axis and sea ice axis. This enables the MPF to be obtained from the polar angle in the transformed coordinate system.

For the application of the polar coordinate transformation, the melt pond and sea ice axes must be correctly determined. As the sea ice axis is poorly defined in Figure 3b and Figure 4, we do not apply the polar coordinate transformation to these two bands, or principal component combinations, directly. Based on the reported spectral albedo values of different sea ice types (Grenfell and Maykut, 1977), we require that the combination of band 2 or blue band (hereinafter referred to as $b2$) and the difference of band 2 and band 8 (hereinafter referred to as $b28$) should theoretically provide the best result (Figure 5). This is because the spectral albedo values of different summer sea ice features differ from each other maximally with this combination (Grenfell and Maykut, 1977). Snow-covered ice and bare ice are separated in $b2$ and in $b28$, which leads to a well-defined linear sea ice axis. $b2$ can also characterize the spectral variation of pixels associated with the transition between 0 and 100% MPF. We have tested bands 3 (green) and 4 (red) as replacements for $b2$; however, $b2$ provides the optimum distinction between the relevant surface classes.

We apply the polar coordinate transformation to $b2$ and $b28$. The 2D scatterplots in cartesian coordinates (Figure 5) show that pixels containing at least a fraction of open lead lie on an axis which has the lowest values of both $b2$ and $b28$. These pixels can be distinguished using simple thresholds and will not be involved in the calculation of MPF. Melt pond pixels lie along the melt pond axis, including melt pond 1 and melt pond 2. Most pixels including bare ice and snow-covered ice lie on the sea ice axis, with a $b2$ value significantly higher than in the melt pond axis. Mixed pixels (melt pond 1 with snow-covered ice and melt pond 2 with snow-covered ice) lie between melt pond and the sea ice axes as expected.

The 2D scatterplot of Figure 5 is transformed to polar coordinates with the sea ice axis and the melt pond axis as the reference axes, with the formulation as follows:

$$\theta = \arctan \left[\frac{b2 - y_0}{b28 - x_0} \right] - \arctan \left[\frac{b2_{mp1} - y_0}{b28_{mp1} - x_0} \right] \quad (1)$$

$$r = \sqrt{(b28 - x_0)^2 + (b2 - y_0)^2} \quad (2)$$

where (x_0, y_0) is the Cartesian coordinate of the center (the intersection of the melt pond axis and sea ice axis), θ is the angle between a pixel's spectral properties ($R2, R2-R8$) and the melt pond axis, r is the radial coordinate, $b2$ is the reflectance of band 2, $b28$ is the reflectance difference between band 2 and band 8, $b2_{mp1}$ is the mean band 2 reflectance of the training area of melt pond 1, $b28_{mp1}$ is the mean $b28$ value of the training area of melt pond 1. The angular coordinate θ is used to determine MPF. r is associated with spectral variation between different sea ice types or melt pond types in the pixel.

To illustrate the transformation, θ and r are plotted in Cartesian coordinates in Figure 6. The sea ice axis and melt pond axis are parallel to each other in this figure, and the two mixed pixel clusters lie between the two. One threshold θ_t is needed to distinguish the pure sea ice pixel from the pixel with MPF above zero, and MPF is proportional to θ as shown in Figure 6. The formulation is as follows:

$$MPF = \begin{cases} 0 & \theta > \theta_t \\ \frac{\theta_t - \theta}{\theta_t - 0.02} & 0.02 < \theta < \theta_t \\ 1 & \theta < 0.02 \end{cases} \quad (3)$$

In this case, the melt pond 1 and melt pond 2 clusters have a calculated MPF value of exactly 100%, and the mixed pixels 1 and mixed pixel 2 clusters which were selected along the edge of ponds have calculated values of 88% and 87%, respectively. This new algorithm is named the ‘LinearPolar’ melt pond retrieval algorithm.

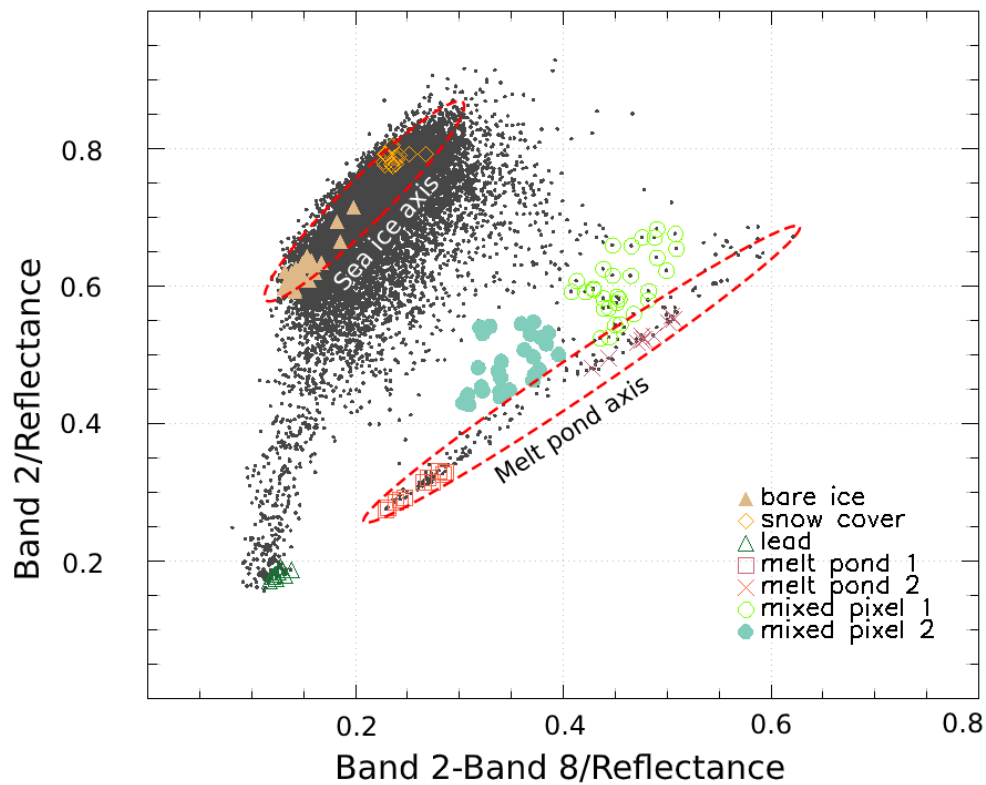


Figure 5. 2-D scatterplot of Sentinel-2 reflectances of b2 vs b28 for the selected subset in Figure 2a. Black dots refer to the coordinate value of unselected pixels in figure.2a)

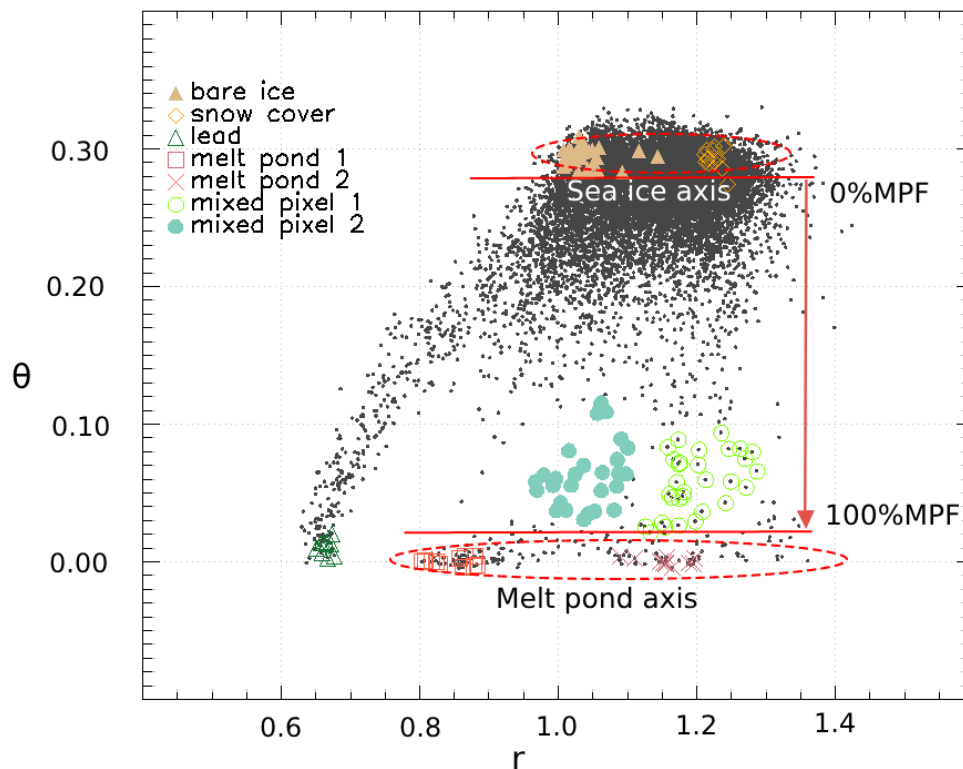


Figure 6. 2-D scatterplot of Sentinel-2 reflectances in the transformed new coordinate system, with axes θ and r (equations 1 and 2). Black dots refer to the coordinate value of unselected pixels in figure 2a)

3.4 Automatic identification of the sea ice and melt pond axes

Like other high-resolution methods, it is crucial to ensure that the 100% melt pond and 100% sea ice axes defined in our algorithm are correct between scenes. We have applied the workflow from Section 3.3 to calculate the fixed pond and sea ice axes from our entire training dataset, and then we applied these fixed axes to all Sentinel-2 images. However, the selected training areas are still not necessarily representative of all image scenes covering all possible sea ice melting conditions. Varying illumination conditions/month/latitude or ice thickness/type can all contribute to the uncertainty of the axes selection. To account for these expected variations, we have developed a workflow to automatically identify these axes using the Hough Transform method (Duda et al., 1972) which enables the algorithm to be applied to all scenes.

The 2-D pixel density scatterplots were segmented into a sea ice cluster and melt pond cluster using a fixed division line (Figure 7a). As the number of pixels in the melt pond class is usually much smaller than in the sea ice class, the calculation can be overly biased towards the sea ice class; therefore, the two axes must be calculated separately. The (θ, ρ) parameter space (Figure 7b) of the two segmented clusters are computed using the Standard

Hough Transform. The variable ρ is the distance from the origin to a certain line of the image domain (Figure 7a) along a vector perpendicular to the line, θ is the angle of the perpendicular projection from the origin to the line measured in degrees clockwise from the positive x -axis. (Duda et al., 1972). A line in the image domain (Figure 7a) is represented by a single point in (θ, ρ) parameter space, with the most likely straight line in the image domain expressed as:

$$\rho = x \cos\theta + y \sin\theta \quad (4)$$

Here, θ and ρ are obtained from the highest density locations of the Hough transform (black squares in Figure 7b) where the sinusoids intersect within the parameter space. When θ and ρ are determined, Equation 4 is the function of the optimized (most probable) straight line through the data.

The location of each axis changes between images. The straight-line parameter (θ, ρ) varied by $(-63 \pm 5.2, 11 \pm 2.7)$ and $(-45 \pm 6.1, -13 \pm 3.3)$ between the 40 Sentinel-2 images we used (Figure 1), for melt pond and sea ice axes, respectively. As shown in Figure 7c, the most likely straight lines identified by the Hough transformation accurately locate the sea ice and melt pond axes.

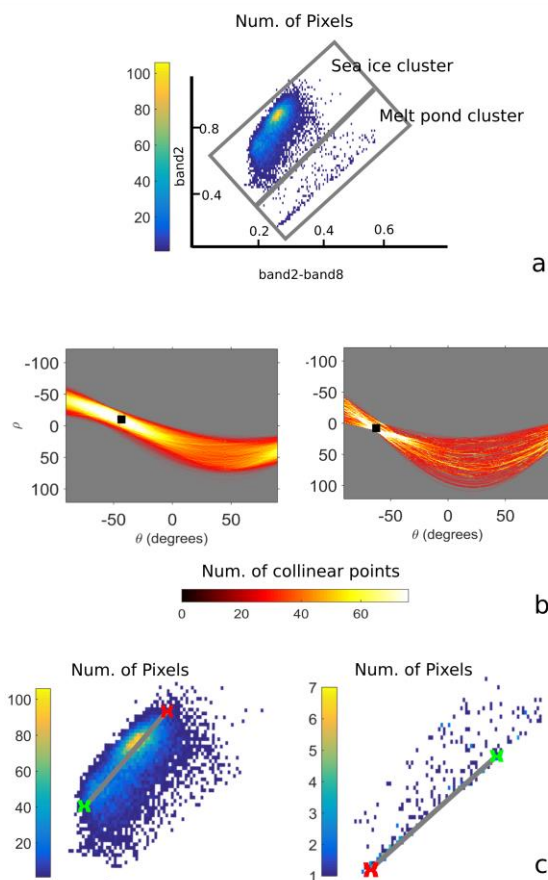


Figure 7. (a) The segmented sea ice cluster and melt pond cluster of the 2-D pixel density scatterplots in Figure 5. (b) Standard Hough Transform of the segmented two clusters, the

color indicates the density of curve intersections. (c) Automatic identification results of the sea ice axis and the melt pond axis (gray line and crosses).

4. Result

4.1 Comparison with PCA and Markus algorithms

In section 3, we demonstrated that the former two algorithms underestimate the pond fraction when ponds are less saturated in the blue channel (training area melt pond 2). We compare the results of each of the three algorithms applied to the same training area, including the LinearPolar method, in Figure 8. The LinearPolar algorithm gives a more accurate MPF for manually assigned 100% ponded pixels. All retrieval results in this section are determined with consistent parameters, having been trained from all the Sentinel-2 images used in this study (i.e., the mean reflectance from Table 1).

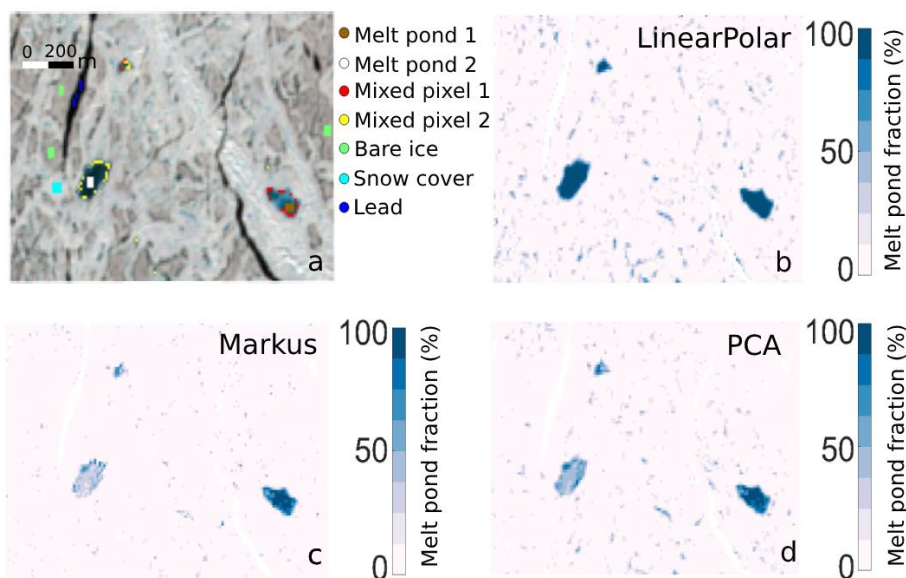


Figure 8. (a) True-color (band combination 4-3-2) subset from the Sentinel-2 scene displayed in Figure 2a with selected training areas. (b) MPF result using the LinearPolar algorithm. (c) MPF result using the Markus algorithm. (d) MPF result using the PCA method. The large size melt pond on the right of the image is melt pond 1, the large one on the left is melt pond 2.

Figure 9 gives the true color images of 5 samples with typical melt ponds, of the total 40 samples included in the 5 scenes shown in Figure 1. These 5 typical samples are divided into two types of melt pond by the color saturation. Samples a and c have a higher saturation than samples b, d, e (see the second column of Figure 9). The results from the LinearPolar algorithm and the two previous algorithms indicate that all the three correctly identify the MPF of class 1 melt ponds, although the Markus and PCA algorithms still appear to underestimate it slightly (samples a, c in Figure 9). However, only the LinearPolar algorithm correctly calculates the MPF of class 2 melt ponds (samples b, d, e in Figure 9), while the

two previous algorithms underestimate it by about 50% on average. These two algorithms also underestimate the MPF of some mixed pixels which include class 2 melt ponds.

We selected 40 samples from the 5 scenes in Figure 1, and the MPF retrieval results of these samples are listed in Table 2. The mean MPF from the LinearPolar algorithm is 8.0%, compared to a mean of 4.9% from PCA and 4.1% from the Markus algorithm. The LinearPolar algorithm produces the highest MPF on average, and also in every sample. The 40 selected Sentinel-2 samples have an MPF from the LinearPolar algorithm 95% and 63% higher than MPF determined from Markus and PCA algorithms, respectively.

4.2 Sensitivity test of threshold

All these algorithms require a threshold input to discriminate 100% sea ice-covered pixels. The sensitivity of the algorithm to the threshold value is a major issue, because the threshold uncertainty can be the main limitation of the accuracy of these algorithms (Rösel and Kalescheke, 2011). The appropriate threshold can change with the image conditions of different scenes, which depends on such factors as solar zenith angle, time of day, season, and cloud cover. If an algorithm has lower sensitivity to threshold variation, then it is more robust. The threshold is expected to vary by ± 0.01 based on statistics of the 40 samples we documented in Table 1. Therefore, we investigated the sensitivity of these algorithms to an expected threshold variation of $\theta_t \pm 3\%$ in each of our 40 samples (Table S1). The average MPF of all samples determined with the LinearPolar algorithm varied by 9.64%, while the method of Markus varied by 20.17% and PCA algorithm varied by 14.12% (Table 2). Thus, the LinearPolar algorithm has the lowest sensitivity to threshold variation among these algorithms and is approximately half as sensitive to its subjective threshold than the previous algorithms.

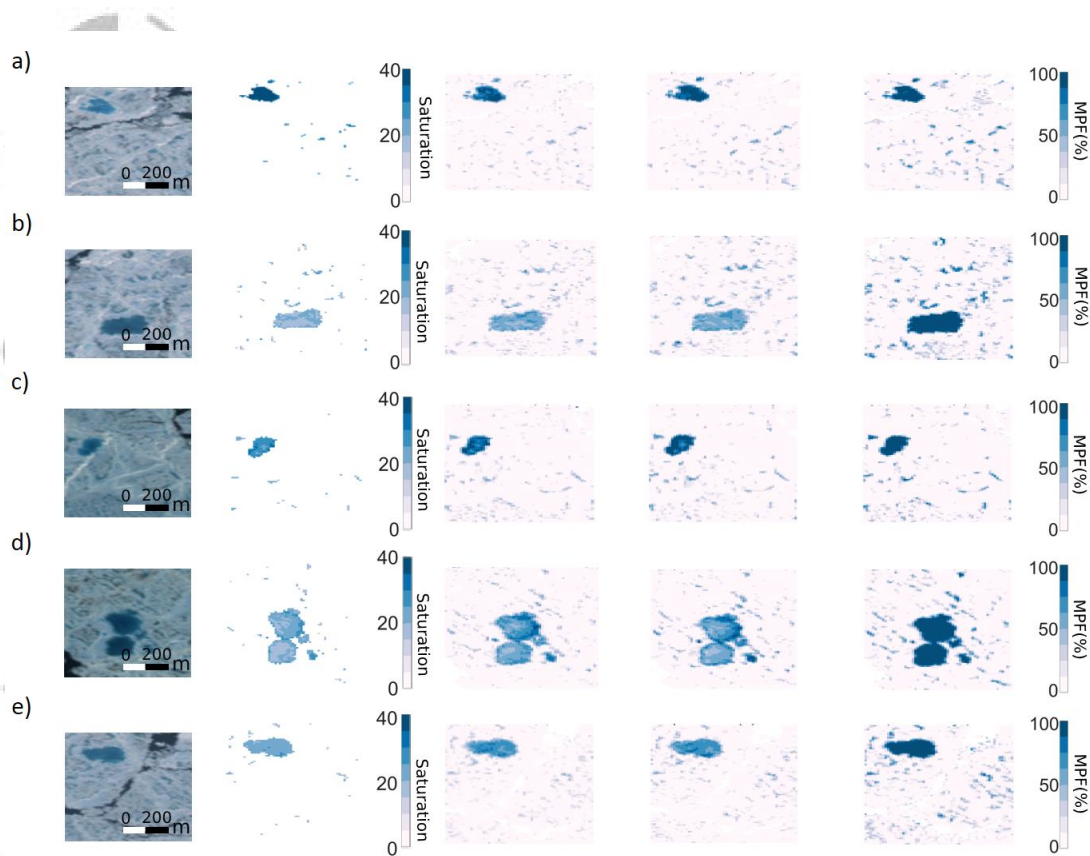


Figure 9. (a)-(e) are 5 selected samples among the 40 samples. The columns from left to right are: 1) True color images, 2) Color saturation of melt pond, 3) MPF determination result of Markus algorithm (%), 4) MPF determination result of PCA algorithm (%), and 5) MPF determination result of LinearPolar algorithm (%). Leads (open water) are masked with white.

Table 2. Comparison of MPF (%) results and MPF variation (%) with a threshold variation of $\theta_t \pm 3\%$ for all 40 training areas.

Method	MPF			MPF Variance		
	Markus	PCA	LinearPolar	Markus	PCA	LinearPolar
Mean	4.10±2.96	4.92±3.73	8.00±6.89	20.17±7.04	14.1±5.16	9.64±4.24

4.3 Comparison with IceBridge dataset

Retrieval results from the three algorithms are assessed using the DMS aerial photographs from Operation IceBridge. The DMS flight shown in Figure 1 is the only flight we found that is coincident with the Sentinel-2 track on the same day. We selected 35 photographs from the flight with clear melt ponds. The melt pond fraction of the 35 photographs used for verification ranged from 10% to 32%. We ran the LinearPolar algorithm on all Sentinel-2 pixels coinciding with DMS images and found that 95% of pixels were classified as having mixed classes (i.e. MPF below 100%). Therefore, the comparison results in this section generally evaluate the impact of mixed pixels on the efficacy of our algorithm. A

Support Vector Machine (SVM) method (Corinna and Vapnik, 1995) is used to distribute the aerial images into 4 classes: melt pond, ice, lead, and ridged ice shadow. The ice shadow class is added because many of the areas in the images are highly ridged and the solar inclination when these images were acquired was very low, which leads to a considerable fraction of ridge shadow. These shadow pixels have an RGB spectrum similar to melt ponds and therefore tend to be classified as melt pond pixels, if they are not explicitly separated. We do not allow aerial image pixels to have mixed classes, i.e. MPF below 100%, because the resolution is 0.1m.

The SVM method is one of the most well-used classification algorithms based on machine learning. Here, all three bands of the aerial images are used. Given the order-of-magnitude higher resolution of these aerial images compared to the Sentinel-2 data, and with appropriately selected training areas, we found that the SVM classification result produced a reliable reference dataset of melt pond fraction. Classification results from the aerial images were then projected to the same UTM projection of Sentinel-2 images for inter-comparison, and MPF is sampled from the average pond fraction within every Sentinel-2 10-m pixel. Pond fractions were determined from the Sentinel-2 image coincident to the aerial photos using each of the three algorithms with the default parameters from the mean reflectance in Table 1.

The LinearPolar algorithm produced an overall melt pond fraction closest to the aerial observations; the estimation error was reduced by a factor of 2 from 40% for the two previous algorithms to only 20% for LinearPolar (Table 3). The RMSE value of the LinearPolar algorithm was about 30% lower than in the previous algorithms. The MPF retrieval results of the four different methods along the aerial observation track are shown in Figure 10. Retrieval results across three different sea ice regimes along the DMS aerial observation track are shown in Figure 11 and Figure S2, S3.

Example 1 is from a region of highly ridged sea ice, including a large number of small but bright blue melt ponds in the low-lying area between ice ridges (Figure 11). In the aerial image with a resolution of 0.1 m, these melt ponds can be easily discriminated, but in a 10-m resolution satellite image such as Sentinel-2, the signal of these bright blue melt ponds can be partially masked by the high reflectivity of the adjacent ridged ice. This leads to an underestimation of MPF. As we have discussed in section 4, all three algorithms have a similar ability to discriminate bright blue melt ponds, so this underestimation is explained by mixing of pond and ice pixels masking the true MPF. Example 2 represents an area where the sea ice is mainly covered by dark blue melt ponds, well developed over smooth ice (Figure S2). The MPF results, in this case, prove our discussion in section 4: the LinearPolar algorithm improves the accuracy of MPF retrieval over dark blue ponds compared to previous methods. Example 3 is from an area mainly covered by bright blue melt ponds, and in comparison to Example 1, the sea ice surface is relatively smooth which has enabled several large ponds to form. All three algorithms can discriminate these melt ponds accurately and give an MPF similar to the aerial image results, although the LinearPolar algorithm is closest (Figure S3).

Table 3. Average MPF (%) results of the 3 algorithms and their Average Error (AE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Relative Error (RE), as compared with IceBridge result.

Method	IceBridge	LinearPolar	PCA	Markus
Average MPF	11.61%	9.18%	6.93%	6.73%
AE		3.19%	5.01%	6.40%
MAE		3.99%	6.24%	6.40%
RMSE		4.69%	6.78%	6.75%
RE		29%	46%	46%

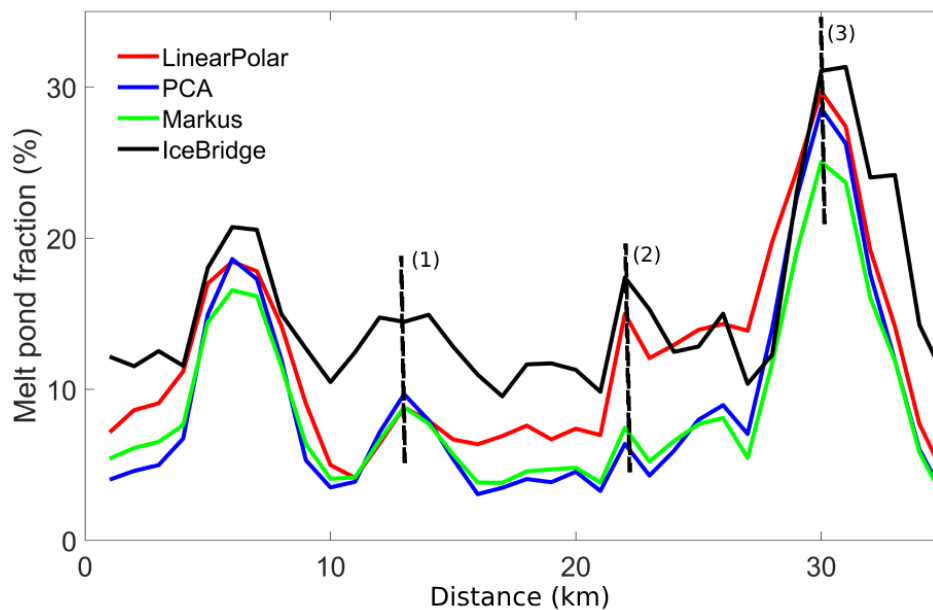


Figure 10. Melt pond fraction determined by the three algorithms along the IceBridge aerial observation flight track (Red, Blue and Green line). Melt pond fraction discrimination result of the IceBridge aerial image (Blackline). Subset 1 is shown in Figure 11; subsets 2 and 3 are shown in Figures S2 and S3.

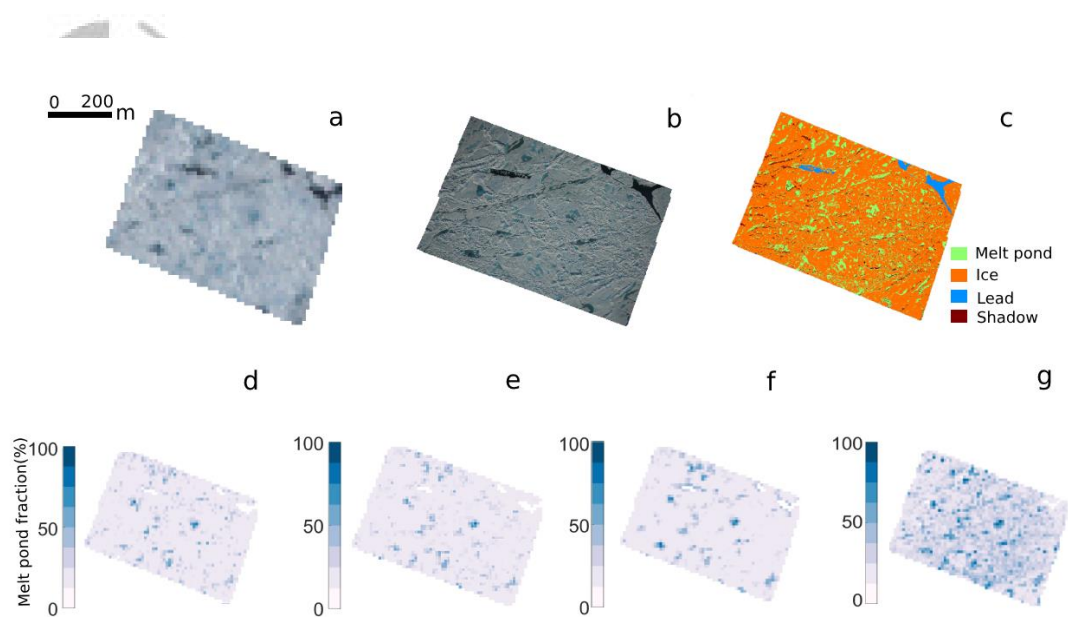


Figure 11. a) True color images of Sentinel-2, b) IceBridge aerial image, c) classification results of IceBridge aerial images, d) MPF result of Markus algorithm (%), e) MPF result of PCA algorithm (%), and f) MPF result of LinearPolar algorithm (%), g) MPF of IceBridge aerial images

5. Discussion

The LinearPolar algorithm is a function that remaps cartesian images to polar space, a technique usually exploited in computer vision. In this paper, we applied this approach to a melt pond retrieval algorithm applicable to high-resolution visible satellite imagery. The spectral reflectance of melt ponds varies over a wide range (broadband reflectance 0.1 to 0.7), acting as the main source of error for determining MPF in previous algorithms. This error has been lowered in medium-resolution imagery, with a physically-based algorithm using radiative transfer modeling to aid spectral unmixing of ponds with different reflectance properties from white ice in medium-resolution imagery (MERIS) (Zege et al., 2015). In this study, we proposed a new algorithm to improve the error which applies to high-resolution (10s of meters) sensor. In the LinearPolar algorithm, we allow melt pond spectral reflectance to vary, enabling 100% MPF over ponds with different spectral properties, which removes this limitation. The method does not rely on assumptions for ice and pond scattering properties like the Zege algorithm does, while still accounting for variable pond reflectance.

By analyzing image samples over ponds with varying properties (owing to, for example, different illumination conditions, pond depth, and ice thickness), the LinearPolar algorithm is capable of accurate MPF retrievals for pixels with 0% -100% pond coverage. We note that, the algorithms based on high-resolution optical satellite imagery used in this study do not treat a pixel as composed of ponded ice and open water. These pixels form a small

quantity in our study because Sentinel-2 has a high resolution of 15 m, and the size of a sea ice floe is usually orders-of-magnitude larger than that. In a broken ice field, there are some small ice floes, but there are seldom melt ponds left on these floes because meltwater can more easily drain off the edges of small floes. Moreover, the melt ponds contribute far less to albedo-related ice thermodynamics in these areas compared to the total open water fraction.

The percentage of a pixel covered by ice versus pond could be accurately determined by the threshold, or it could be biased by the wrong threshold because the reflectance properties of ice and pond within the pixel are not known perfectly a priori. This has important implications for melt pond fraction retrievals from lower-resolution optical imagery, e.g., 250-m MODIS pixels, where pixels almost always contain a mixture of pond and ice/snow or lead classes. A workflow based on the Standard Hough Transform method could be applied to determine the correct threshold a priori for separating class fractions within mixed pixels.

By comparing the MPF retrieval results of the three algorithms with in-situ aerial observations, we have shown that the LinearPolar algorithm improves the accuracy of retrievals over dark blue ponds. Since a poor detection ability of bright blue melt ponds in highly ridged sea ice regions (with shadows) is found in all three algorithms, MPF determination algorithms should be applied carefully in such regions.

6. Conclusion

In this study, we have used the Sentinel-2 sensor to determine MPF. Our results prove that the former algorithms of Markus and PCA can effectively distinguish melt ponds with high color saturation (bright blue) but they largely underestimate melt ponds with lower saturation (darker blue). We propose a new algorithm, named the LinearPolar algorithm, to correct this problem. The LinearPolar technique accounts for melt pond spectral reflectance variation by determining an axis of 100% pond coverage across the full range of pond color saturation levels (i.e., bright to dark blue). The Hough Transform method can be used to automatically identify the sea ice and melt pond axes, thereby improving the automatic applicability of the algorithm. A fixed reflectance threshold is still required to separate areas of zero MPF (i.e., bare ice or snow) from mixed pixels with some fraction covered by ponds and 100% pond pixels. However, the new algorithm can effectively group melt ponds with variable reflectance properties and is 52% and 32% less sensitive to its semi-subjective threshold selection than the Markus and PCA techniques, respectively. We have demonstrated that the LinearPolar algorithm results are more precise and accurate than previous methods.

The mean MPF determined with the Markus, PCA and LinearPolar algorithms for our 40 selected Sentinel-2 subsamples are 4.1%, 4.9% and 8.0% respectively. Owing to the improved discrimination of low saturation melt ponds, pond fractions retrieved from the LinearPolar algorithm are 95 % and 63% higher than, respectively, those from the Markus and PCA algorithms. According to the assessment using the DMS aerial photographs from

Operation IceBridge, the RMSE value of the LinearPolar algorithm is about 30% lower than in the previous algorithms.

Acknowledgments

This research was supported by the National Key R&D Program of China (No.2018YFA0605903), and the National Key Research and Development Program of China (2016YFC1402704). We are thankful for the Sentinel-2 data provided by the European Space Agency (<https://scihub.copernicus.eu/>), and the IceBridge DMS images provided by the National Snow and Ice Data Center (<https://nsidc.org/icebridge/portal/map>).

References

1. Bindschadler R., Vornberger P., Fleming A., et al. (2008). The Landsat Image Mosaic of Antarctica. *Remote Sensing of Environment*, 112(12):4214-4226.
2. Curry J. A., Schramm J. L., Ebert E. E. (1995). Sea Ice-Albedo Climate Feedback Mechanism. *Journal of Climate*, 8(2):240-247.
3. Cortes C., Vapnik V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273-297.
4. Drusch M., Bello U. D., Carlier S., et al. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Proceedings of SPIE - The International Society for Optical Engineering*, 120(none):0-36.
5. Duda R. O., Hart P. E., (1972). Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11-15.
6. E. Zege, A. Malinka, I. Katsev, A. Prikhach, G. Heygster, L. Istomina, G. Birnbaum, p. Schwarz, Algorithm to retrieve the melt pond fraction and the spectral albedo of Arctic summer ice from satellite optical data, *Remote Sensing of Environment*, Volume 163, 2015, Page 153-164.
7. Grenfell T. C., Maykut G. A., (1977). The Optical Properties of Ice and Snow in the Arctic Basin. *Journal of Glaciology*, 18(80):445-463.
8. Hanebeck, U. D., (2001). Template matching using fast normalized cross correlation. *Proceeding of Spie on Optical Pattern Recognition XII*, 4387, 95-102.
9. Istomina L, Heygster G, Huntemann M, et al. Melt pond fraction and spectral sea ice albedo retrieval from MERIS data – Part 1: Validation against in situ, aerial, and ship cruise data[J]. *The Cryosphere*, 2015, 9(4):1551-1566.
10. Istomina L, Heygster G, Huntemann M, et al. Melt pond fraction and spectral sea ice albedo retrieval from MERIS data – Part 2: Case studies and trends of sea ice albedo and melt ponds in the Arctic for years 2002–2011[J]. *The Cryosphere*, 9,4(2015-08-12), 2015, 9(4):1567-1578.
11. Istomina L, Melsheimer C, Huntemann M, et al. Retrieval of sea ice thickness during melt season from in situ, airborne and satellite imagery[C]// *IGARSS 2016 - 2016 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2016.

12. Landy J. C., Ehn J. K., Shields M., et al. (2015). Surface and melt pond evolution on landfast first-year sea ice in the Canadian Arctic Archipelago. *Journal of Geophysical Research: Oceans*, 119(5): 3054-3075.
13. Landy J. C., Ehn J. K., Barber D. G. (2015). Albedo feedback enhanced by smoother Arctic sea ice. *Geophysical Research Letters*, 42(24):10714-10720
14. Lu, P., Leppäranta, M., Cheng, B., Li, Z., Istomina, L., Heygster, G. (2018). The color of melt ponds on Arctic sea ice. *The Cryosphere*, 12(4): 1331–1345.
15. Lu, P., Cao, X., Wang, Q., Leppäranta, M., Cheng, B., Li, Z. (2018). Impact of a thin ice lid on the optical properties of melt ponds. *Journal of Geophysical Research: Oceans*, 123(11): 8313-8328.
16. Lu P, Li Z, Cheng B, et al. Sea ice surface features in Arctic summer 2008: Aerial observations[J]. *Remote Sensing of Environment*, 2010, 114(4):693-699.
17. Li H, Perrie W, Li Q, et al. Estimation of Melt Pond Fractions on First Year Sea Ice Using Compact Polarization SAR[J]. *Journal of Geophysical Research: Oceans*, 2017, 122(10):8145-8166.
18. Malinka A, Zege E, Heygster G, et al. Reflective properties of white sea ice and snow[J]. *Cryosphere*, 2016, 10(6):2541-2557.
19. Markus T., D.J. Cavalieri and A. Ivanoff. (2002). The potential of using Landsat 7 ETM+ for the classification of sea-ice surface conditions during summer., *Annals of Glaciology*, 34, 415–419.
20. Malinka A, Zege E, Istomina L, et al. Reflective properties of melt ponds on sea ice[J]. *The Cryosphere*, 2018, 12(6):1921-1937.
21. Makynen M, Kern S, Rösel A, et al. On the Estimation of Melt Pond Fraction on the Arctic Sea Ice With ENVISAT WSM Images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(11):7366-7379.
NASA Distributed Active Archive Center, National Snow and Ice Data Center, Boulder, CO. (2012). IceBridge DMS L1B geolocated and orthorectified images (IODMS1B, 2012). Retrieved from <http://nsidc.org/data/iodms1b/html>
22. Morassutti M. P., Ledrew E. F. (2015). Albedo and depth of melt ponds on sea ice. *International Journal of Climatology*, 16(7):817-838.
23. Palmer M. A., Saenz B. T., Arrigo K. R. (2014). Impacts of sea ice retreat, thinning, and melt-pond proliferation on the summer phytoplankton bloom in the Chukchi Sea, Arctic Ocean. *Deep-Sea Research Part II*, 105(7):85-104.
24. Perovich D. K., Iii W. B. T., Ligett K. A. (2002). Aerial observations of the evolution of ice surface conditions during summer. *Journal of Geophysical Research Oceans*, 107(C10): SHE-24-1-SHE 24-14.
25. Perovich, D. K. (1996). 96-1 The Optical Properties of Sea Ice, CRREL Monogr., vol. 96-1, 25 pp., U.S. Office of Naval Res., Springfield, Va.
26. Paul V. C. Hough, (1961). "Method and means for recognizing complex patterns", U.S. Patent 3069654,
27. Polashenski C, Perovich D, Courville Z. The mechanisms of sea ice melt pond formation and evolution[J]. *Journal of Geophysical Research: Oceans*, 2012, 117(C1).

28. Rösel, A., Kaleschke L. Birnbaum G. (2012). Melt ponds on Arctic sea ice determined from MODIS satellite data using an artificial neural network. *The Cryosphere*, 6, 2(2012-04-03), 6(2):431-446.
29. Rösel A., Kaleschke L. (2011). Comparison of different retrieval techniques for melt ponds on Arctic sea ice from Landsat and MODIS satellite data. *Biometrical Journal*, 47(3):299–308.
30. Scharien R K, Landy J, Barber D G . First-year sea ice melt pond fraction estimation from dual-polarisation C-band SAR; Part 1: In situ observations[J]. *The Cryosphere*, 2014a, 8(6):2147-2162.
31. Scharien, R.K., Hochheim, K., Landy, J. and Barber, D.G., 2014b. 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(6), pp.2163-2176.
32. Taylor, P. D., Feltham, D. L. (2004). A model of melt pond evolution on sea ice. *Journal of Geophysical Research: Oceans*, 109(C12).
33. Tanaka, Y., Tateyama, K., Kameda, T., and Hutchings, J. K. (2016), Estimation of melt pond fraction over high-concentration Arctic sea ice using AMSR-E passive microwave data, *J. Geophys. Res. Oceans*, 121, 7056– 7072, doi:10.1002/2016JC011876.
34. Tschudi M. A., Maslanik J. A., Perovich D. K., (2008). Derivation of melt pond coverage on Arctic sea ice using MODIS observations. *Remote Sensing of Environment*, 112(5):2605-2614.
35. Tschudi M. A., Curry J. A., Maslanik J. A. (2001). Airborne observations of summertime surface features and their effect on surface albedo during FIRE/SHEBA. *Journal of Geophysical Research Atmospheres*, 106(D14):15335-15344.
36. Yackel, J. J. , & Barber, D. G. . (2000). Melt ponds on sea ice in the canadian archipelago: 2. on the use of radarsat-1 synthetic aperture radar for geophysical inversion. *Journal of Geophysical Research: Oceans*, 105(C9), 22061.

Accept