Open Source Algorithm for Detecting Sea Ice Surface Features in High Resolution Optical Imagery

Nicholas C. Wright[1], Christopher M. Polashenski[1,2]

[1]Thayer School of Engineering, Dartmouth College, Hanover, NH, USA

[2]U.S. Army Cold Regions Research and Engineering Laboratories, Hanover, NH, USA

Correspondence to: N. C. Wright (ncwright.th@dartmouth.edu)

Abstract. Snow, ice, and melt ponds cover the surface of the Arctic Ocean in fractions that change throughout the seasons. These surfaces control albedo and exert tremendous influence over the energy balance in the Arctic. Increasingly available m- to dm-scale resolution optical imagery captures the evolution of the ice and ocean surface state visually, but methods for quantifying coverage of key surface types from raw imagery are not yet well established. Here we present an open source system designed to provide a standardized, automated, and reproducible technique for processing optical imagery of sea ice. The method classifies surface coverage into three main categories: Snow and bare ice, melt ponds and submerged ice, and open water. The method is demonstrated on imagery from four sensor platforms and on imagery spanning from spring thaw to fall freeze-up. Tests show the classification accuracy of this method typically exceeds 96%. To facilitate scientific use, we evaluate the minimum observation area required for reporting a representative sample of surface coverage. We provide an open source distribution of this algorithm and associated training data sets and suggest the community consider this a step towards standardizing optical sea ice imagery processing. We hope to encourage future collaborative efforts to improve the code base and to analyze large datasets of optical sea ice imagery.

1 Introduction

The surface of the sea ice-ocean system exhibits many different forms. Snow, ice, ocean, and melt ponds cover the surface in fractions that change throughout the seasons. The relative fractions of these surfaces covering the Arctic ocean are undergoing substantial change due to rapid loss of sea ice (Stroeve et al., 2012), increase in the duration of melt (Markus et al., 2009; Stroeve et al., 2014), decrease in sea ice age (Maslanik et al., 2011), and decrease in sea ice thickness (Kwok and Rothrock, 2009; Laxon et al., 2013) over recent decades. As a whole, the changes are reducing albedo and enhancing the absorption of solar radiation, triggering an ice albedo feedback (Curry et al., 1995; Perovich et al., 2008; Pistone et al., 2014). Large-scale remote sensing has been instrumental in documenting the ongoing change in ice extent (Parkinson and Comiso, 2013), thickness (Kurtz et al., 2013; Kwok and Rothrock, 2009; Laxon et al., 2013), and surface melt state (Markus et al., 2009). An increasing focus on improving prediction of future sea ice and climate states, however, has also created substantial interest in better observing, characterizing, and modeling the *processes* that drive changes in albedo-relevant sea ice surface conditions such as melt pond formation, which occur at smaller length scales. For these, observations that resolve surface conditions explicitly are needed to understand the underlying causes of the seasonal and spatial evolution of albedo in a more sophisticated way.

34   Explicitly sensing the key aspects of the sea ice surface, including melt pond coverage, degree of deformation, floe

35   size, and lead distributions, requires evaluating the surface at meter to decimeter scale resolution. Variability in the

36   spatial coverage and morphology of these surface characteristics, however, occurs over hundreds of meters to tens of

37   kilometers. Estimates of aggregate scale surface coverage fraction must therefore be made at high resolution over

38   sample domains of many square kilometers. Quantifying the relative abundance of surface types over domains of

39   multi-kilometer scale from manned ground campaigns is both time consuming and impractical. Remote sensing

40   provides a more viable approach for studying these multi-kilometer areas. High resolution optical imagery (e.g. Figure

41   1) visually captures the surface features of interest, but the methods for analyzing this imagery remain under-

42   developed.

43   The need for remote sensing methods enabling quantification of meter-scale sea ice surface characteristics has

44   been well recognized, and efforts have been made to address it. Recent developments in remote sensing of sea ice

45   surface conditions fall into two categories: (1) methods using low-medium resolution satellite imagery (i.e. having

46   pixel sizes larger than the typical ice surface feature size) with spectral un-mixing type algorithms to derive aggregate

47   measures of sub-pixel phenomena (e.g. for melt ponds Markus et al., 2003; Rösel et al., 2012; Rösel and Kaleschke,

48   2011; Tschudi et al., 2008) and (2) methods using higher resolution satellite or airborne imagery (i.e. having pixel size

49   smaller than the typical scale of ice surface features) that is capable of explicitly resolving features (e.g. Arntsen et

50   al., 2015; Fetterer and Untersteiner, 1998; Inoue et al., 2008; Kwok, 2014; Lu et al., 2010; Miao et al., 2015; Perovich

51   et al., 2002b; Renner et al., 2014; Webster et al., 2015). The first category, those derived from low-medium resolution

52   imagery, have notable strengths in their frequent sampling and basin-wide coverage. They cannot, however, provide

53   detailed statistics on the morphology of surface features necessary for assessing our process-based understanding and

54   have substantial uncertainty due to ambiguity in spectral signal un-mixing. The second category – observations at high

55   resolutions which explicitly resolve surface properties – can provide these detailed statistics, but were historically

56   limited by a dearth of data acquisitions. Recent increases in imagery availability from formerly classified defense

57   (Kwok, 2014) or commercial satellites (e.g. DigitalGlobe), and increases in manned flights over the Arctic (e.g.

58   IceBridge, SIZRS) have substantially reduced this constraint for optical imagery. Likely increases in collection of

59   imagery from UAV's (DeMott and Hill, 2016) and increases in satellite imaging bandwidth (e.g. DigitalGlobe

60   WorldView 4 launched in 2016) suggest that availability of high resolution imagery will continue to increase.

61   Processing high resolution sea ice imagery to derive useful metrics quantifying surface state, however, remains a

62   major hurdle. Recent years have seen numerous publications demonstrating the success of various processing

63   techniques for optical imagery of sea ice on limited test cases (e.g. Inoue et al., 2008; Kwok, 2014; Lu et al., 2010;

64   Miao et al., 2015; Perovich et al., 2002b; Renner et al., 2014; Webster et al., 2015). None of these techniques, however,

65   have been adopted as a standard or been used to produce large-scale datasets, and validation has been limited.

66   Furthermore, none have been challenged by imagery collected across the seasonal evolution of the ice or used to

67   process data from multiple sensor platforms. These issues must be addressed to enable in large scale production-type

68   image processing and use of high resolution imagery as a sea ice monitoring tool.

69   A unique aspect of high resolution sea ice imagery datasets, which differs from most satellite remote sensing, is

70   the quantity of image sources and data owners. Distributed collection and data ownership means centralized processing

71   of imagery to produce a single product is unlikely. Instead, we believe that distributed processing by dataset owners

72   is more likely and the community therefore has a substantial need for a shared, standard processing protocol.

73   Successful creation of such a processing protocol would increase imagery analysis and result in the production of

74   datasets suitable for ingestion by models to validate surface process parameterizations. In this paper, we assess

75   previous publications detailing image processing methods for remote sensing and present a novel scheme that builds

76   from the strengths and lessons of prior efforts. Our resulting algorithm, the Open Source Sea-ice Processing (OSSP)

77   Algorithm, is presented as a step toward addressing the community need for a standardized methodology, and released

78   in an open source implementation for use and improvement by the community.

79   We began with three primary design goals that guided our development of the image processing scheme. The

80   method must (1) have a fully automatic workflow and have a low barrier to entry for new users, (2) produce accurate,

81   consistent results in a standardized output format, and (3) be able to produce equivalent geophysical parameters from

82   a range of disparate image acquisition methods. To meet these goals, we have packaged OSSP in a user-friendly

83   format, with clear documentation for start-up. We include a set of default parameters that should meet most user needs,

84   permitting processing of pre-defined image types with minimal set-up. The algorithm parameters are tunable to allow

85   more advanced users to tailor the method to their specific imagery input. We chose an open source format to enhance

86   the ability for the community to explore and improve the code relative to a commercial software. Herein, we discuss

87   how we arrived at the particular technique we use, and why it is superior to some other possible mechanisms. We then

88   demonstrate the ability of this algorithm to analyze imagery of disparate sources by showing results from high

89   resolution DigitalGlobe WorldView satellite imagery in both panchromatic and pansharpened formats, aerial sRGB

90   (standard Red, Green, Blue) imagery, and NASA Operation IceBridge DMS (Digital Mapping System) optical

91   imagery. In this paper, we classify imaged areas into three surface types: Snow and ice, melt ponds and submerged

92   ice, and open water. The algorithm is, however, suitable for classifying any number of categories, should a user be

93   interested in different surface types, and might be adapted for use on imagery of other surface types.

94   **2 Algorithm Design**

95   Two core decisions were faced in the design of this image classification scheme: (1) Whether to analyze the image by

96   individual pixels or to analyze objects constructed of similar, neighboring pixels, and (2) which algorithm to use for

97   the classification of these image units.

98   Prior work in terrestrial remote sensing applications has shown that object-based classifications are more accurate

99   than single pixel classifications when analyzing high-resolution imagery (Blaschke, 2010; Blaschke et al., 2014; Duro

100   et al., 2012; Yan et al., 2006). In this case, 'high resolution' has a specific definition dependent on the relationship

101   between the size of pixels and objects of interest. An image is high resolution when surface features of interest are

102   substantially larger than pixel resolution and therefore are composed of many pixels. In such imagery, objects, or

103   groups of pixels constructed to contain only similar pixels (i.e. a single surface type), can be analyzed as a set. The m-

104   dm resolution imagery meets this definition for features like melt ponds and ice floes. Object based classification

105   enables an algorithm to extract information about image texture and spatial correlation within the pixel group;

106   information that is not available in single pixel based classifications and can enhance accuracy of surface type

107   discrimination. Furthermore, object based classifications are much better at preserving the size and shape of surface

108   cover regions. Classification errors of individual pixel schemes tend to produce a 'speckled' appearance in the image

109   classification with incorrect pixels scattered across the image. Errors in object based classifications, meanwhile,

110   appear as entire objects that are mislabeled (Duro et al., 2012). Since our intent is to process high-resolution imagery

111   and produce measurements not only of the areal fractions of surface type regions, but also to enable analysis of the

112   size and shape of ice surface type regions (e.g. for floe size or melt pond size determination), the choice of object

113   based classification over pixel based was clear.

114   A wide range of algorithms were considered for classifying image objects. We first considered the use of

115   supervised versus an unsupervised classification schemes. Unsupervised schemes were rejected as they produce

116   inconsistent, non-intercomparable results. These schemes, such as clustering algorithms, group observations into a

117   predefined number of categories – even if not all feature types of interest are present in an image. For example, an

118   image containing only snow-covered ice will still be categorized into the same number of classes as an image with

119   snow, melt ponds, and open water together – resulting in multiple classes of snow. Since the boundary between classes

120   also changes in each image, standardizing results across imagery with different sources and of scenes with different

121   feature content would be challenging at best.

122   Supervised classification schemes instead utilize a set of known examples (called training data) to assign a

123   classification to unknown objects based on similarity to user-identified objects. Supervised classification schemes

124   have several advantages. They can produce fixed surface type definitions, allow for more control and fine tuning of

125   the algorithm, improve in skill as more points are added to the training data, and allow users to choose what surface

126   characteristics they wish to classify. While many machine learning techniques have shown high accuracy in remote

127   sensing applications (Duro et al., 2012), we selected a random forest machine learning classifier over other supervised

128   learning algorithms for its ability to handle nonlinear and categorical training inputs (Breiman, 2001; DeFries, 2000;

129   Pal, 2005), resistance to outliers in the training dataset (Breiman, 1996), and relative ease of implementation.

130   Our scheme, learning from the success of Miao et al. (2015) in classifying aerial imagery, uses an image

131   segmentation algorithm to divide the image into objects which are then classified with random forest machine learning.

132   Our implementation of the segmentation and classification, however, were custom-built using well known image

133   processing tools (Pedregosa et al., 2011; van der Walt et al., 2014) in an open source format. We do not attempt to

134   assert that our method is the optimal method for processing sea ice imagery. Instead, we argue that it is easily usable

135   by the community at large, produces highly accurate and consistent results, and merits consideration as a standardized

136   methodology. In coordination with this publication, we release our code (available at https://github.com/wrightni/ossp

137   doi:10.5281/zenodo.1133689) with the intention of encouraging movement toward a standardized method. Our hope

138   is to continue development of the algorithm with contributions and suggestions from the sea ice community.

3 Methods

3.1 Image Collection and Preprocessing

The imagery used to test the algorithm was selected from four distinct sources in order to assess the algorithm's ability to deliver consistent and intercomparable measures of geophysical parameters. We chose high resolution satellite imagery from DigitalGlobe's WorldView constellation in panchromatic and 8 band multispectral formats, NASA Operation IceBridge Digital Mapping System optical imagery, and aerial sRGB imagery collected using an aircraft-mounted standard DLSR camera as part of the SIZONet project. We first demonstrate the technique's ability to handle imagery representing all stages of the seasonal evolution of sea ice conditions on a series of 22 panchromatic satellite images collected between March and August of 2014 at a single site in the Beaufort Sea: 72.0° N 128.0° W. We then process 4 multispectral WorldView 2 images of the same site, each collected coincident with a panchromatic image and compare results to assess the benefit of spectral information. Finally, we process a set of 20 sRGB images and 20 IceBridge DMS images containing a variety of sea ice surface types to illustrate the accuracy of the method on aerial image sources. The imagery sources chosen for this analysis were selected to be representative of the variation that exists in optical imagery of sea ice, but there is an abundance of image data that can be processed with this technique.

The satellite images were collected by tasking WorldView 1 and WorldView 2 Digital Globe satellites over fixed locations in the Arctic. Tasking requests were submitted to DigitalGlobe with the support and collaboration of the Polar Geospatial Center. The panchromatic bands of WorldView 1 and 2 both have a spatial resolution of 0.46m at nadir. The WorldView 1 satellite panchromatic band samples the visible spectrum between 400 nm and 900 nm, while the WorldView 2 satellite panchromatic band samples between 450 nm and 850 nm. In addition, WorldView 2 has 8 multispectral bands at 1.84 m nadir resolution, capturing bands within the range of 400nm to 1040nm. Each WorldView image captures an area of ~700-1300 km^2. Of the 22 useable panchromatic collections at the site, 15 were completely cloud free while 7 of the images were partially cloudy. Images with partial cloud cover were manually masked and cloud covered areas were excluded from analysis. The aerial sRGB imagery was captured along a 100km long transect to the north of Barrow, Alaska with a Nikon D70 DSLR mounted at nadir to a light airplane during June 2009. The IceBridge imagery was collected in July of 2016 near 73° N 171° W with a Canon EOS 5D Mark II digital camera. We utilize the L0 (raw) DMS IceBridge imagery, which has a 10cm spatial resolution when taken from 1500 feet altitude (Dominguez, 2010, updated 2017).

Each satellite image was orthorectified to mean sea level before further processing. Orthorectification corrects for image distortions caused by off-nadir acquisition angles and produces a planimetrically correct image that can be accurately measured for distance and area. Due to the relatively low surface roughness of both multiyear and first year sea ice (Petty et al., 2016), errors induced by ignoring the real topography during orthorectification are small. Multispectral imagery was pansharpened to the resolution of the panchromatic imagery. Pansharpening is a method that creates a high resolution multispectral image by combining intensity values from a higher resolution panchromatic image with color information from a lower resolution multispectral image. The pansharpened imagery used here was created using a 'weighted' Brovey algorithm. This algorithm resamples the multispectral image to the resolution of the panchromatic image, then each pixel's value is multiplied by the ratio of the corresponding panchromatic pixel

175 value to the sum of all multispectral pixel values. The orthorectification and pansharpening scripts were developed by

176 the Polar Geospatial Center at the University of Minnesota and utilize the GDAL (Geospatial Data Abstraction

177 Library) image processing tools (GDAL, 2016). All imagery used was rescaled to the full 8-bit color space for

178 improved contrast and viewing. No other preprocessing was done to the aerial sRGB imagery or IceBridge DMS

179 imagery.

3.2 Image Segmentation

181 A flow chart of the image processing steps taken after pre-processing is presented in Fig. 2. The first task in the image

182 processing algorithm is to segment the image into groups of similar pixels, called objects. Accurate segmentation

183 requires finding the boundaries between the natural surface types we wish to differentiate (e.g. the boundary between

184 ice covered and open ocean), delineating their locations, and using these boundaries to produce image objects. Sea ice

185 surface types have large differences in reflectivity and tend to change abruptly, rather than gradually over a large

186 distance. We exploit this characteristic by using an edge detection algorithm to find boundaries between surface types.

187 Figure 3 contains a visual demonstration of this process. First, a Sobel-Feldman operator (van der Walt et al., 2014)

188 is applied to the input image (Fig. 3a). The Sobel-Feldman filter applies a discrete differentiation kernel across the

189 image to find the local gradient of the image intensity. High gradient values correspond to abrupt changes in pixel

190 intensity, which are likely boundaries between surface types. We scale the gradient values by an amplification factor

191 of 2 in order to further highlight edge regions in the image. Following the amplification, we threshold the lowest 10%

192 of the gradient image and set the values to zero. This reduces noise detected by the Sobel-Feldman filter, and eliminates

193 weaker edges. The amplification factor and gradient threshold percentage are both tuning parameters, which can be

194 adjusted to properly segment images based on the input image and the strength of edges sought.

195   The strongest edges in optical imagery of sea ice are typically the ocean-ice interface, followed by melt pond-ice

196 boundaries, then ice ridges and uneven ice surfaces. In general, the more edges detected, the more segmented the

197 image will become, and the more computational resources required to later classify the increased number of image

198 objects. On the other hand, an under-segmented image may miss the natural boundaries between surfaces. Under

199 segmentation introduces classification error because an object containing two surface types cannot be correctly

200 classified. An optimally segmented image is one which captures all the natural surface boundaries with minimal over-

201 segmentation (i.e. boundaries placed in the middle of features). The appropriate parameters for our imagery were

202 tuned by visual inspection of the segmentation results. In such inspection, desired segmentation lines are manually

203 drawn, and algorithm-determined segmentation lines are overlain and evaluated for completeness.

204   The result of the edge detection is a gradient map that marks the strength of edges in the image. We use a watershed

205 segmentation technique to build complete objects based on edge locations and intensity (van der Walt et al., 2014).

206 We first calculate all local minimum values in the gradient image, where a marker is then placed to indicate the origin

207 of watershed regions. Each region then begins iteratively expanding in all directions of increasing image gradient until

208 encountering a local maximum in the gradient image or encountering a separately growing region. This continues until

209 every pixel in the image belongs to a unique set. With the proper parameter selection, each object will represent a

210 single surface type. It is often the case that some areas will be over-segmented (i.e. a single surface feature represented

211 by multiple objects). Over segmentation can either be ignored, or objects can be recombined if they meet similarity

212 criteria in an effort to save computational resources. Here we chose to classify objects without recombination. Figure

213 3b shows the detected edges overlain on top of the input image.

214   The watershed segmentation algorithm benefits from the ability to create objects of variable size. Large objects

215 are built in areas of low surface variability while many small objects are created in areas of high variability. This

216 variable object sizing is well suited to sea ice surface classification because the variability of each surface type occurs

217 at different scales. Areas of open water and snow covered first year ice, for example, can often be found in large

218 expanses, while areas that contain melt ponds, ice ridges, or rubble fields frequently cover small areas and are tightly

219 intermingled with other surface types. Variable object sizes give the fine detail needed to capture surfaces of high

220 heterogeneity in their full detail, while limiting over segmentation of uniform areas.

221 **3.3 Segment Classification**

222 **3.3.1 Overview**

223 Once the image has been divided into regions of the same surface type, each object must be classified as to which

224 surface type it represents. We classify the objects using a random forest machine learning technique (Breiman, 2001;

225 Pedregosa et al., 2011). The development of a machine learning algorithm requires multiple iterative steps: 1) Select

226 attributes with which to classify each object, 2) create a training dataset, 3) classify unknown image objects based on

227 the training set, and 4) assess performance and refine, starting from step 1. Random forest classifiers excel for their

228 relative ease of use, flexibility in the choice of attributes that define each object, and overall high accuracy even with

229 relatively small training datasets. The random forest classifier is only one of many available machine learning

230 approaches and others may also be suitable.

231 **3.3.2 Surface Type Definitions**

232 Another key challenge to quantitatively monitoring sea ice surface characteristics from high resolution imagery is a

233 lack of standardized surface type definitions. We noted above that high-resolution sea ice imagery comes from many

234 sources; each with different characteristics. As we will see below, each image source will need to have its own training

235 set created by expert human classifiers. The human classifier must train the algorithm according to definitions of each

236 surface type that are broadly agreed upon in the community for the algorithm to be successful in producing

237 intercomparable datasets. While at first the definitions of open water, ice and melt ponds might seem intuitive, many

238 experts in the cryosphere community have differing opinions, especially on transitional states. Deciding where to

239 delineate transitional states is important to standardization. We have established the following definitions for the three

240 surface types we sought to separate, binning transitional states in a manner most consistent with their impact on albedo.

241 Our surface type definitions focus on the behavior of a surface in absorption of shortwave radiation and radiative

242 energy transfer. (1) Open Water (OW): Applied to surface areas that had zero ice cover as well as those covered by

243 an unconsolidated frazil or grease ice. (2) Melt Ponds and Submerged Ice (MPS): Applied to surfaces where a liquid

244 water layer completely submerges the ice. (3) Ice and Snow: Applied to all surfaces covered by snow or bare ice, as

245 well as decaying ice and snow that is saturated, but not submerged. The definition of melt ponds includes the classical

246 definition of melt ponds where meltwater is trapped in isolated patches atop ice, as well as optically-similar ice
247 submerged near the edge of a floe. While previous work separates these categories (e.g. (Miao et al., 2015) we did not
248 attempt to break these 'pond' types because the distinction is unimportant from a shortwave energy balance (albedo)
249 perspective. We further refined the ice and snow category into two sub categories: (3a) Thick Ice and Snow, applied
250 during the freezing season to ice appearing to the expert classifier to be thicker than 50cm or having an optically thick
251 snow cover and to ice during the melt season covered by a drained surface scattering layer (Perovich, 2005) of
252 decaying ice crystals and (3b) Dark and Thin Ice, applied during the freezing season to surfaces of thin ice that are not
253 snow covered including nilas and young ice. This label was also applied during melting conditions to ice covered by
254 saturated slush, but not completely submerged in water. This is ice which in some prior publications (e.g. Polashenski
255 et al., 2012) was labeled as 'slushy bare ice'. We acknowledge that the boundary between the ice and snow sub-
256 categories is often more a continuum than a defined border but note that distinguishing the two types is useful for
257 algorithm accuracy. Dividing the ice/snow type creates two relatively homogeneous categories rather than a single
258 larger category with large internal differences. A user only interested in the categories of ice, ponds, and open water
259 could simply re-combine them, as we have done for analysis. We created a temporary 'shadow' classification category
260 that was used only in panchromatic WorldView images captured prior to melt onset. This category allowed the
261 machine learning algorithm to differentiate dark shadows in spring imagery from melt ponds in summer imagery –
262 surface types that look similar in single-band imagery. This category is used exclusively as a temporary step in
263 processing that allows us to bypass masking shadow regions directly (e.g. (Webster et al., 2015). As this is not
264 designed to be a standalone classification category (as opposed to Miao et al., 2015, 2016), objects classified as a
265 shadow were grouped back with the ice/snow category.

266 **3.3.3 Attribute Selection**

267 Attributes are quantifiable measures of image object properties used by the classifier in discriminating surface types.
268 An enormous array of possible attributes could be calculated for each image object and could be calculated in many
269 ways. Examples of properties that could be quantified as attributes include values of the enclosed pixels, the size and
270 shape of the object, and values of adjacent pixels. The calculation of pixel values aggregated by image objects takes
271 advantage of the additional information held in the pixel group (as compared to individual pixels). We have compiled
272 a list representing a relevant subset of such attributes that can be used to distinguish different surface types in Table
273 1. We included a selection of attributes similar to those used in previous publications (e.g. Miao et al., 2015), as well
274 as attributes we have developed specifically for our algorithm.
275 Each image source provides unique information about the surface and it can be expected that a different list of
276 attributes will be optimal for classification of each image type – even though we seek the same geophysical parameters.
277 As high-resolution satellite images can have millions of image objects, calculating the attributes of each object quickly
278 becomes computationally expensive. We have, therefore, determined those that are most valuable for classifying each
279 image type to use in our classification. For example, pansharpened WorldView 2 imagery has 8 spectral bands which
280 can inform the classification, while panchromatic versions of the same image have only a single band. Our goal was
281 to select a combination of attributes that describe the intensity and textural characteristics of the object itself, and of

282 the area surrounding the object. Table 1 indicates which attributes were selected for use in classifying each image
283 type.

284   We selected attributes by only including those with a high relative importance. The importance of each attribute
285 is a property of a random forest classifier, and is defined as the number of times a given attribute contributed to the
286 final prediction of an input. After initial tests with large numbers of attributes, we narrowed our selection by using
287 only those attributes that contributed to a classification in greater than 1% of cases. For discussion here, we group the
288 attributes into two broad categories: Those calculated using internal pixels alone (object attributes), and those
289 calculated from external pixel values (neighbor attributes).

290 3.3.4 Object Attributes

291 The most important attributes in the classification of an image segment were found to be aggregate measures of pixel
292 intensity within the object. We determine these by analyzing the mean pixel intensity of all bands and the median of
293 the panchromatic band. An important benefit of image segmentation is the ability to calculate estimates of surface
294 texture by looking at the variability within a group of pixels. The texture is often unique in the different surface types
295 we seek to distinguish. Open water is typically uniformly absorptive and has minimal intensity variance. Melt ponds,
296 in contrast, come in many realizations and exhibit a wider range in reflectance, even within individual ponds. To
297 estimate surface texture, we calculate the standard deviation of pixel intensity values and the image entropy within
298 each segment. Image entropy, H, is calculated as

$$299 \qquad H = -\sum p * \log_2 p$$

300 where p represents the bin counts of a pixel intensity histogram within the segment. We also calculate the size of each
301 segment as the number of pixels it contains. As sea ice surface characteristics evolve appreciably over time,
302 particularly before and after melt onset, we use image acquisition date as an attribute in for classification. While date
303 of melt onset varies, and the reader might argue that a more applicable attribute would be image melt state, melt state,
304 is not an apriori characteristic of the image. It would therefore need to be manually defined, and not meet our demand
305 for a fully automated scheme.

306   In multispectral imagery, we also calculate the ratios between the mean absorption of each object in certain
307 portions of the spectrum. The important band ratios used for the multispectral WorldView imagery were determined
308 empirically. We tested every possible band combination, and successively removed the ratios that did not contribute
309 to more than 1% of object classifications. In sRGB imagery we use the band ratios shown to be informative in this
310 application by Miao et al. (2015).

311   In addition to information contained within each object, we utilize information from the surrounding area. To
312 analyze the surrounding region, we determine the dimensions of a minimum bounding box that contains the object,
313 then expand the box by five pixels in each direction. All pixels contained within this box, minus those in the object,
314 are considered to be neighboring pixels. Analogous to the internal attribute calculations, we find the average intensity
315 and standard deviation of these pixels. We also calculate the maximum single intensity within the neighboring region.
316 Searching for attributes outside of the object improves the algorithm's predictive capabilities by providing spatial
317 context. Bright neighboring pixels (as an analog for an illuminated ridge) often provide information to distinguish, for

318   example, a shadowed ice surface from a melt pond. In panchromatic imagery, melt ponds and shadows appear similar

319   when evaluated solely on internal object attributes. However, a dark region with an immediately adjacent bright region

320   is more likely to be a shadow than a dark region not adjacent to a bright pixel (e.g. a pond). We do note that it is likely

321   that a more complex algorithm, for example identifying those pixels in a radius or distance to the edge of the segment,

322   rather than using a bounding box, would be more reliable. The tradeoff, however, is one of higher computational

323   expense.

324   **3.4 Training Set Creation**

325   Four training datasets were created to analyze the images selected for this paper. One training set was created for

326   each imagery source: Panchromatic satellite imagery, multispectral satellite imagery, aerial sRGB imagery, and

327   IceBridge DMS imagery. Each training set consists of a list of image objects that have been manually classified by a

328   human and a list of attribute values calculated from those objects and their surroundings. The manual classification is

329   carried out by multiple sea ice experts. Experienced observers of sea ice can classify the majority (85%+) of segments

330   in a high resolution optical image with confidence. To address the ambiguity in correct identification of certain

331   segments, however, we used several (4) skilled sea ice observers to repeatedly classify image objects. For the initial

332   creation of our training datasets, two of the users had extensive training in the OSSP algorithm and surface type

333   definitions, while the other two no experience with the algorithm. Users in both categories were briefed on the standard

334   surface type definitions used for this study (section 3.3.2). Figure 4 shows a confusion matrix to compare user

335   classifications. Cells in the diagonal indicate agreement between users, while off- diagonal cells indicate disagreement

336   (Pedregosa et al., 2011). Agreement between the two well-trained users was high (average 94% of segment

337   identifications; Fig. 4a), while the agreement between a well-trained user and a new user was lower (average of 86%;

338   Fig 4b). After an in-person review of the training objects among all four users, the overall agreement rose to 97%. The

339   remaining 3% of objects were cases where the expert users could not agree on a single classification, even after review

340   of the surface type definitions and discussion. These objects were therefore not used in the final training set. Figure 5

341   shows a series of surface types that span all our classification categories, including those where the classification is

342   clear and those where it is difficult. Difficult segments are over-represented in these images for illustrative purposes,

343   and represent a relatively small fraction of the total surface.

344   While the skill of the machine learning prediction increases substantially as the size of the training set grows,

345   creating large training sets is time consuming. We found that training datasets of approximately 1000 points yielded

346   accurate and consistent results. We have developed a graphical user interface (GUI) to facilitate the rapid creation of

347   large training sets (see Fig. 6). The GUI presents a user with the original image side by side with an overlay of a single

348   segment on that image. The user assigns a classification to the segment by visual determination.

349   The training dataset is a critical component of our algorithm because it directly controls the accuracy of the

350   machine learning algorithm – and using a consistent training set is necessary for producing intercomparable results.

351   In coordination with this publication we are releasing our version 1.0 training datasets with the intention that they

352   would represent a first version of *the* standard training set to use with each image type. Though we have found this

353   training dataset robust through our error analyses below, it is our intention to solicit broader input from the community

354   to refine and expand the training datasets available and release future improved versions.

355   In addition to cross-validating the creation of a training dataset between users, we assess the quality of our training

356   set through an out-of-bag (OOB) estimate, which is an internal measure of the training set's predictive power. The

357   random forest method creates an ensemble (forest) of classification trees from the input training set. Each classification

358   tree in this forest is built using a random bootstrap sample of the data in the training set. Because training samples are

359   selected at random, each tree is built with an incomplete set of the original data. For every sample in the original

360   training set, there then exists a subset of classifiers that do not contain that sample. The error rate of each classifier

361   when used to predict the samples that were left out is called the OOB estimate (Breiman, 2001). The OOB estimate

362   has been shown to be equivalent to predicting a separate set of features and comparing the output to a known

363   classification (Breiman, 1996).

364   **3.5 Assigning Classifications**

365   Once the training dataset is complete, the algorithm is prepared to predict the classification of unknown objects in the

366   images. The random forest classifier is run and a classified image is created by replacing the values within each

367   segment by the classification label predicted. Figure 3c shows the result of labeling image objects with their predicted

368   classification. From the classified image, it is possible to produce a number of useful statistics. The most basic

369   measurement is the total pixel counts for each of the three surface categories. This provides both the total area, in

370   square kilometers, that each surface covers, and the fraction of each image that is covered by each surface type. It

371   would also be possible to calculate measurements such as the average segment size for each surface, melt pond size

372   and connectivity, or floe size distributions. Each of these, however, has its own standardization problems significant

373   enough to merit their own paper.

374   For demonstration, we have used the output from our image classification to calculate the fractional melt pond

375   coverage for each date. The melt pond fraction was defined as the area of melt ponds and submerged ice divided by

376   the total area covered by ice floes, i.e.:

377 $$Melt\ Pond\ Coverage = \frac{Area_{MPS}}{Area_{MPS} + Area_{I+S}}$$

378   where the subscript MPS indicates predicted melt ponds and submerged ice and I+S indicates predicted ice and snow.

379   **3.6 Determining Classification Accuracy**

380   The primary measure of classification accuracy was to test the processed imagery on a per pixel basis against human

381   classification. For every processed image, we selected a simple random sample of 100 pixels chosen from the whole

382   image and asked four sea ice experts to assign a classification to those pixels. For a single image from each image

383   source we also asked the sea ice experts to classify and additional 900 pixels. This larger sample was created to

384   demonstrate a tighter confidence interval, while the smaller samples were chosen to demonstrate consistency across

385   images. We used the same GUI developed to create training datasets to assess pixel accuracy. Pixels were presented

386   at random to the user by showing the original image with the given pixel highlighted. The user then identified which

387   of the surface type categories best described that pixel. This assignment is then compared to the algorithm's prediction

388   behind the scenes. The accuracy, as determined by each of the four experts, was averaged to create a composite

389   accuracy for each image.

390   **4 Results**

391   **4.1 Classification of Four Imagery Sources**

392   The OSSP image processing method proved highly suitable for the task of classifying sea ice imagery. A visual

393   comparison between the raw and processed imagery, shown in Fig. 7 can quickly demonstrate this in a qualitative

394   sense. Figure 7 contains a comparison between the original and classified imagery for each source, selected to show

395   the performance of the algorithm on images that contain a variety of surface types. The colors shown correspond to

396   the classification category; regions colored black are open water, blue regions are melt ponds and submerged ice, gray

397   regions are wet and thin ice, and white regions are snow and ice. The quantitative processing results, including surface

398   distributions and classification accuracy are shown in Table 2. The overall classification accuracy was $96 \pm 3\%$ across

399   20 IceBridge DMS images; $95 \pm 3\%$ across 20 aerial sRGB images; $97 \pm 2\%$ across 22 panchromatic WorldView 1

400   and 2 images; and $98 \pm 2\%$ across 4 multispectral WorldView 2 images.

401   The nature of the classification error is presented using a confusion matrix that compares the algorithm

402   classification with a manual classification for 1000 randomly selected pixels. Four confusion matrices, one for a single

403   image from each of the four image sources is shown in Fig. 8. Values along the diagonal of the square are the

404   classifications where the algorithm and the human observer agreed, while values in off-diagonal areas indicate

405   disagreement. Concentration of error into a particular off-diagonal cell helps illustrate the types of confusion the

406   algorithm experiences. The number of pixels that fall into off-diagonal cells is low across all imagery types. In the

407   IceBridge imagery, there is a slight tendency for the algorithm to classify surfaces as open water where a human would

408   choose melt pond. This is caused by exceptionally dark melt ponds on the edge of melting through (Fig. 5, panels F

409   and I). Classification of mutlispectral WorldView imagery has a small bias towards classifying melt ponds over dark

410   or thin ice (Fig. 5, panel D). Aerial sRGB and Panchromatic WorldView images do not have a distinct pattern to their

411   classification errors.

412   The internal metric of classification training dataset strength, the Out of Bag Error (OOB) estimates, on a 0.0 to

413   1.0 scale, are shown in Table 3 for the trees built from our three training sets. The OOB estimate represents the mean

414   prediction error of the random forest classifier, i.e. an OOB score of 0.92 estimates that the decision tree would predict

415   92% of segments that are contained in the training dataset correctly. The discrepancy between OOB error and the

416   overall classification accuracy is a result of more frequent misclassification of smaller objects; overall accuracy is area

417   weighted, while the OOB score is not.

418   **4.2 WorldView: Analyzing A Full Seasonal Progression**

419   We analyzed 22 images at a single site in the Beaufort Sea collected between March and August of 2014 to challenge

420   the method with images that span the seasonal evolution of ice surface conditions. The site is Eulerian; it observes a

421   single location in space rather than following a single ice floe through its lifecycle as it drifts. Still, the results of these
422   image classifications (shown in Fig. 9) illustrate the progression of the ice surface conditions in terms of our four
423   categories over the course of a single melt season. While cloud cover impacted the temporal continuity of satellite
424   images collected at this site, we are still able to follow the seasonal evolution of surface features. A time series of
425   fractional melt pond coverage calculated from the satellite image site is plotted in Fig. 10. The melt pond coverage
426   jumps to 22% in the earliest June image, as initial ponding begins and floods the surface of the level first year ice.
427   This is followed by a further increase to 45% coverage in the next few days. The melt pond coverage then drops back
428   down to 30% as melt water drains from the surface and forms well defined ponds. The evolution of melt pond coverage
429   over our satellite observation period is consistent with prior field observations (Eicken, 2002; Landy et al., 2014;
430   Polashenski et al., 2012) and matches the four stages of ice melt first described by Eicken (2002). The ice at this
431   observation site fully transitions to open water by mid-July, though it appears that the ice is advected out of the region
432   in the late stages of melt rather than completing melt at this location.

433   **5 Discussion**

434   **5.1 Error**

435   There are four primary sources of error in the OSSP method as presented, two internal to the method and two external.
436   Internal error is caused by segment misclassification and by incomplete segmentation (i.e. leaving pixels representing
437   two surface types within one segment). The net internal error was quantified in section 3.6 and 4. External error is
438   introduced by pixilation – or blurring of real surface boundaries due to insufficient image resolution – and human
439   error in assigning a 'ground truth' value to an aerial or satellite observation during training.

440   **5.1.1 Internal Error**

441   Through assessing the accuracy of each classified image on a pixel-by-pixel basis (section 3.6), we collect all internal
442   sources of error into one measurement: The algorithm either assigned the same classification as a human would have,
443   or it did not. Total internal accuracy calculated for the method, relative to human classifiers, is quite good, at 90-99%
444   across all image types. Our experience is that this level of accuracy approaches the accuracy with which fractional
445   surface coverage can practically be determined from labor intensive ground campaign techniques such as lidar and
446   measured linear transects (e.g. Polashenski et al., 2012)
447   The first type of internal error is misclassification error, where the image classification algorithm fails to assign
448   the same classification that a human expert would choose. This type of error is best quantified by analyzing the training
449   datasets. The OOB score for each forest of decision trees (Table 3) provides an estimate of each forest's ability to
450   correctly predict objects similar to those used to create the forest (section 3.4). The OOB score is not influenced by
451   segmentation error, because the objects selected for training dataset use were filtered to remove any objects that
452   contained more than one surface type. The most commonly misapplied category was the Dark and Thin Ice
453   subcategory of Ice and Snow. This category often represents surface types that are in a transitional state, and is often
454   difficult to classify even for a human observer.

The second type of internal error is segmentation error, where an object is created that contains more than one of the surface types we are trying to distinguish. This occurs when boundaries between objects are not placed where boundaries between surfaces exist; an issue most common where one surface type gradually transitions to another. When this occurs, some portion of that object will necessarily be misclassified. We have compensated for areas that lack sharp boundaries by biasing the image segmentation towards over-segmentation, but a small number of objects still contain more than one surface type. During training set creation, we asked the human experts to identify objects containing more than one surface type. 3.5% of objects were identified as insufficiently segmented in aerial imagery, and 2% of objects in satellite imagery. This represents the upper limit for the total percentage of insufficiently segmented objects for several reasons. First, segmentation error was most prevalent in transitional surface types (i.e. Dark and Thin Ice), which represents a small portion of the overall image and is composed of relatively small objects. This category is overrepresented in the training objects because objects were chosen to sample each surface type and not weighted by area. In addition, insufficiently segmented objects are generally composed of only two surface types, and end up identified as the surface which represents more of the object's area. Hence the total internal error introduced by segmentation error is appreciably smaller than misclassification error, likely well under 1%.

5.1.2 External Error

The first form of external error is introduced by image resolution. At lower image resolutions, more pixels of the image span edges, and smaller features are more likely to go undetected. Pixels on the edge of surface types necessarily represent more than one surface type, but can be classified as only one. Misclassification of these has the potential to become a systemic error if edge pixels were preferentially placed in a particular category. We assessed this error's impact by taking high resolution IceBridge imagery (0.1m), downsampling to progressively lower resolution, and reprocessing. Figure 11 shows the surface type percentages for three IceBridge images at decreasing resolution. Figure 12 shows a series of downsampled images and their classified counterparts. Surprisingly, despite clear pixilation and aliasing in the imagery, little change in aggregate classification statistics occurred as resolution was lowered from 0.1 to 2m. This suggests that at resolutions used for this paper, edge pixels do not significantly impact the classification results. It may also be possible to forego the pansharpening process discussed in section 3.1, and use 2m multispectral WorldView imagery directly.

The second type of external error occurs when the human expert fails to correctly label a segment. Even skilled human observers cannot classify every pixel in the imagery definitively, and indeed the division between the surface types can sometimes be indistinct even to an observer on the ground. We addressed this concern by employing observers extensively trained in the sea ice field, both in remote sensing and in-situ observations, comparing multiple human classifications of the same segments. After discussion, the portion of image objects subject to human observer disagreement or uncertainty is small. Human observers disagreed on 3% of objects creating our training sets. The possibility of systemic bias among the expert observer classifications cannot be excluded because real ground truth, in the form of geo-referenced ground observations from knowledgeable observers was, unfortunately, not available for any of the imagery. Conducting this type of validation would be helpful, but given high confidence human expert classifiers expressed in their classifications and low disagreement between them, may not be essential.

5.1.3 Overall Error

The fact that misclassification dominates the internal error metric suggests that error could be reduced if additional object attributes used by human experts to differentiate surface types could be identified. The agreement between the OSSP method and a human (96%+/-3%) is similar to the agreement between different human observers (97%), meaning that the algorithm is nearly as accurate as a human manually classifying an entire image. If we exclude the possibility for systemic error in human classification, and assume other errors are unrelated to one another, we can calculate a total absolute accuracy in surface type determination as approximately 96%.

5.2 Producing Derived Metrics of Surface Coverage

The classified imagery, presented as a raster, (e.g. Fig. 7) is not likely to be the end product used in many analyses. Metrics of the sea ice state in simpler form will be calculated. We already introduced the most basic summary metrics in section 4, where we presented fractional surface coverage calculated from the total pixel counts for each of the four surface categories in each image. We also presented the calculation of melt pond coverage as a fraction of the ice-covered portion of the image, rather than total image area. The calculation of these is straightforward. Other metrics commonly discussed in the literature that could be produced with minimal additional processing include those capturing melt pond size, connectivity, or fractal dimension, as well as floe size distribution or perimeter to area ratio. As with definitions of surface type, standardizing metrics will be necessary to produce intercomparable results. We discussed the more complex metrics which could be derived from this imagery with several other groups. We determined that standardizing these and other more advanced metrics will require more input and consensus building before a community standard can be suggested. We leave determining standard methods for calculating these more complex metrics to a future work.

Equipped with the images processed by OSSP, we consider what size area must be imaged, classified, and summarized to constitute 'one observation' and how regionally representative such an observation is. Even with the increasing availability of high resolution imagery, it is unlikely that high resolution imaging will regularly cover more than a small portion of the Arctic in the near future. As a result, high resolution image analysis will likely remain a 'sampling' technique. Since the scale of sea ice heterogeneity varies for each property type, a minimum area unique to that property must be analyzed to qualify as a representative sample of the surface conditions. Finding that minimum area involves addressing the 'aggregate scale' – the area over which a measured surface characteristic becomes uniform and captures a representative average of the property in the area (Perovich, 2005). It may also be possible to determine an aggregate scale statistic within well constrained bounds by random sub-sampling of the region, and therefore reduce processing time. Here we conduct analysis of these sampling concepts and suggest this analysis of the aggregate scale be conducted for any metric.

First, we sought to determine the aggregate scale for the simple fractional coverage metrics of ice as a fraction of total area and melt pond as a fraction of ice area. This would inform us, for example, as to whether processing the entire area of a WorldView image (up to 1000km^2) was necessary, or alternatively if a full WorldView image was sufficient to constitute a sample. First, we evaluated the convergence of fractional coverage within areas of increasing size towards the image mean. For a WorldView image depicting primarily first year ice in various stages of melt, we

527 created non-overlapping gridded subsections and determined the fractional coverage within each grid cell. The size of

528 grid cells was varied logarithmically from 100 x 100 pixels (10^2) to 31622x31622 pixels ($10^{4.5}$) or from 0.0025km^2 to

529 250km^2. For each sample size, we gridded the image and evaluated every subsection within the entire image. Figure

530 13a shows a scatterplot of the fractional melt pond coverage in each image grid plotted against the log of total area of

531 that grid cell. As the area sampled increases, the melt pond fraction shows lower deviation from the mean, as expected.

532 To assist in evaluating the convergence towards the mean, we plot the 95% prediction interval for each image subset

533 size in Fig. 13a (large red dots). The range of pond fraction values between these two points represents the interval

534 within which 95% of samples of this size would fall. The width of the 95% prediction interval declines linearly with

535 respect to sample area in log space, shrinking by 0.3 for each order of magnitude that sample area increases. Visually,

536 it appears that maximum convergence may have been reached at a sample area of ~30km^2 (~$10^{1.5}$km^2), though there

537 are an insufficient number of samples at this large area within a single image to be certain. Regardless of whether

538 convergence is complete, the prediction interval tells us that at 30km^2, 95% of areas sampled could be expected to

539 have pond coverage within 5% of the mean of a full image (~1000km^2). This is consistent with prior work that

540 indicated the aggregate scale for melt pond fraction determination is on the order of several tens of square kilometers

541 (Perovich, 2005; Perovich et al., 2002a). In Fig. 13b we conduct the same analysis for the total ice-covered fraction

542 (ponded + unponded ice) of the image. We see the range of the prediction interval generally drops as larger samples

543 are taken, but does not converge as cleanly or quickly as the pond coverage prediction interval does - a finding that is

544 unsurprising as ice fraction is composed of discrete floes with sizes much larger than melt ponds. The limited

545 convergence indicates that the aggregate scale for determination of ice covered fraction is at least on the order of the

546 scale of a WorldView image, and likely larger. Aggregate scale ice concentration, unlike melt pond fraction, is a

547 statistic better observed with medium resolution remote sensing platforms such as MODIS or Landsat due to the need

548 for a larger satellite footprint. WorldView imagery may be particularly useful for determining smaller scale parts of

549 floe size distributions or for validating larger scale remote sensing of ice fraction, if the larger scale pixels can be

550 completely contained within the worldview image. Floe size distribution will likely require nesting of scales in order

551 to fully access both large and small-scale parts of the floe size distribution.

552   We next investigated whether it is possible to reduce the processing load required to determine the melt pond or

553 ice fraction of an image within certain error bounds by processing collections of random image subsets. To do this, it

554 is useful to first establish two definitions: (1) one random sample of size N represents N randomly selected 100x100

555 pixel boxes, and (2) one adjacent sample of size N is a single area with size $100\sqrt{N}$ x $100\sqrt{N}$. In other words, a

556 random sample and an adjacent sample both represent an image area of 10,000*N pixels, but consist of independent

557 and correlated pixels, respectively. We expect random samples to better represent the total image mean melt pond

558 fraction because ice conditions are spatially correlated and a single large area is not composed of independent samples.

559 We evaluated this hypothesis by collecting 1000 random and adjacent samples of size N=100, with replacement.

560 Results are shown in Fig. 14. In Figure 14a, we plot a histogram of the mean melt pond fraction determined from these

561 1000 samples. The means determined from sets that contained randomly distributed image areas, are in red. The means

562 determined from sets of adjacent image areas are in blue. Although both sets represent samples of the same total image

563 area, the one composed of independent subsets randomly selected from across the image does a much better job of

564 representing the mean value, with a smaller standard deviation.

565 Estimating the mean of a complete image by sampling randomly selected areas of the image becomes a simple

566 statistics problem. The sample size needed to estimate a population mean to within a certain confidence interval and

567 margin of error can be determined with the formula:

568
$$n = \left(\frac{Z\sigma}{ME}\right)^2$$

569 where n is the sample size, Z is the z-score for the confidence interval required, σ is the population standard deviation,

570 and ME is the margin of error. The standard deviation of random samples with size 100 is ~0.2 (Fig 14b). Assuming

571 a 95% confidence interval (Z=1.96), and a margin of error of 5%, a sample size of 64 is required. In other words, with

572 64 random samples of size 100 we can predict the mean melt pond fraction of the total image with 95% confidence to

573 within 5%. 64 samples of size 100 corresponds to an image area of ~16km^2, significantly smaller than the total image

574 size.

575 In order to show these results visually, we return to Fig. 13 and place another set of 95% prediction interval bounds

576 (purple dots). These bounds represent the prediction interval for a random sample of size necessary for the total area

577 to equal the area on the x axis. The result is quite powerful. By processing as little as 16km^2 of the image, collected

578 from samples randomly distributed across the area, we can determine aggregate melt pond fraction to within 5% of

579 the true value with a confidence of 95%. We estimate that this 5% margin of error is comparable to the sum of internal

580 (2-4%) and external errors in our processing algorithm (section 5.1). For large scale processing we suggest that when

581 the sample confidence interval is below the image processing technique accuracy, sampling of larger areas is no longer

582 necessary.

583 A similar analysis is presented in Fig. 14c and 14d for ice fraction. While the WorldView image is likely not large

584 enough to represent the aggregate scale for ice fraction, randomly sampling the image still provides an expedient way

585 to determine the mean ice fraction of the image within certain bounds, while processing only a small fraction of the

586 image. Calculating the 95% prediction interval of random samples representing the total image area shown on the x

587 axis (purple dots) again shows that the total image mean can be estimated by calculating only a small portion of the

588 total image.

589 These explorations of image sampling permit us to recommend that users can estimate the total image pond fraction

590 by selecting N sets of 100 randomly selected 50x50m regions (where N is selected to provide the desired confidence

591 interval and margin of error). We suggest a standard, which incorporates some 'safety factor', for processing imagery

592 to produce estimates of melt pond fraction should be to process 25km^2 of area contained in at least 100 randomly

593 located image subsets from domains of at least 100km^2. We note that flying a UAV over a domain and collecting

594 imagery along flight tracks will not count as fully 'random' in this context, since the images along-track are spatially

595 correlated. Since an image does not represent the aggregate scale for ice fraction, we cannot recommend a specific

596 sampling strategy for the aggregate scale, but note that processing of 25km^2 of imagery from randomly distributed

597 subsets produces a prediction interval around the mean of approximately the same size as the upper limit of uncertainty

598 for our image processing technique. The statistical approach for determining aggregate statistics should not depend

599   on the seasonality of the image nor the type of image used so long as the total area observed is sufficiently greater
600   than the variability in the surface feature being investigated. However, these recommendations should be considered
601   provisional, because they are subject to impacts from differences in ice property correlation scales, and should be
602   further evaluated for accuracy as larger processed datasets are available.

603   **5.3 Community Adoption**

604   We have provided a free distribution of the OSSP algorithm and the training sets discussed in section 3.4 and 4 as a
605   companion to this publication, complete with detailed startup guides and documentation. This OSSP algorithm has
606   been implemented entirely in Python using open source resources with release to additional users in mind. The code,
607   along with documentation, instructional guidelines, and premade training sets (those used for the analyses herein) is
608   available at https://github.com/wrightni/ossp (doi:10.5281/zenodo.1133689). The software is packaged with default
609   parameters and version controlled training sets for 4 different imagery sources. The package includes a graphical user
610   interface to allow users to build custom training datasets that suit their individual needs. The algorithm was constructed
611   with the flexibility to allow for the classification of any number of features given an appropriate training dataset.
612   Our intention is that by providing easy access to the code in an open source format, we will enable both specific
613   inquiries and larger scale image processing that supports community efforts at general sea ice monitoring. We plan to
614   continue improving and updating the code as it gains users and we receive community feedback. We hope to encourage
615   others to design their own features and add-ons. Since the predictive ability of the machine learning algorithm
616   improves as more training data is added, we wish to strongly encourage the use of the GUI to produce additional
617   training sets and we plan to collate other users training sets into improved training versions. See documentation of the
618   training set creation GUI for more information on how to share a training set.
619   The OSSP algorithm helps to bring the goal of having a standardized method for deriving geophysical parameters
620   from high resolution optical sea ice imagery closer to reality. In the larger picture, developing such a tool is only the
621   first step. We recall that the motivation behind this development was the need to quantify sea ice surface conditions
622   in a way that could enable better understanding of the processes driving changes in sea ice cover. The value of the
623   toolkit will only be realized if it is used for these scientific inquiries. We look forward to working with imagery owners
624   to facilitate processing of additional datasets.

625   **6. Conclusions**

626   We have implemented a method for classifying the sea ice surface conditions from high resolution optical imagery of
627   sea ice. We designed the system to have a low barrier to entry, by coding it in an open source format, providing
628   detailed documentation, and releasing it publicly for community use. The code identifies the dominant surface types
629   found in sea ice imagery; open water, melt ponds and submerged ice, and snow and ice, with accuracy that averages
630   96 percent – comparable to the consistency between manual expert human classifications of the imagery. The
631   algorithm is shown to be capable of classifying imagery from a range of image sensing platforms including
632   panchromatic and pansharpened WorldView satellite imagery, aerial sRGB imagery, and optical DMS imagery from

633   NASA IceBridge missions. Furthermore, the software can process imagery collected across the seasonal evolution of

634   the sea ice from early spring through complete ice melt, demonstrating it is robust even as the characteristics of the

635   ice features seasonally evolve. We conclude, based on our error analysis, that this automatic image processing method

636   can be used with confidence in analyzing the melt pond evolution at remote sites.

637   With appropriate processing, high resolution imagery collections should be a powerful tool for standardized and

638   routine observation of sea ice surface characteristics. We hope that providing easy access to the methods and algorithm

639   developed herein, we will facilitate the sea ice community convergence on a standardized method for processing high

640   resolution optical imagery either by adoption of this method, or by suggestion of an alternate method complete with

641   code release and error analysis.

642

643   The authors declare that they have no conflict of interest.

644

645   *Data Availability.* The OSSP algorithm code is available from https://github.com/wrightni/ossp

646   (doi:10.5281/zenodo.1133689). Image data and processing results are available at the NSF Arctic Data Center (ADC).

647   Raw and preprocessed image data from DigitalGlobe WorldView images are not available due to copyright, but can

648   be acquired from DigitalGlobe or the Polar Geospatial Center at the University of Minnesota.

649

650   *Acknowledgements.* This work was supported by the Office of Naval Research Award N0001413MP20144 and the

651   National Science Foundation Award PLR-1417436. We would like to thank Donald Perovich and Alexandra Arntsen

652   for their assistance in creating machine learning training datasets. We would also like to thank Arnold Song, Justin

653   Chen, and Elias Deeb for their assistance and guidance with the development of the OSSP code. WorldView satellite

654   imagery was provided with the DigitalGlobe NextView License through the University of Minnesota Polar Geospatial

655   Center. A collection of the aerial imagery was collected by the SIZONet project. Some data used in this paper were

656   acquired by the NASA Operation IceBridge Project.

657   **References**

658   Arntsen, A. E., Song, A. J., Perovich, D. K. and Richter-Menge, J. A.: Observations of the summer breakup of an

659   Arctic sea ice cover, Geophys. Res. Lett., 42(19), 8057–8063, doi:10.1002/2015GL065224, 2015.

660   Blaschke, T.: Object based image analysis for remote sensing, ISPRS J. Photogramm. Remote Sens., 65(1), 2–16,

661   doi:10.1016/j.isprsjprs.2009.06.004, 2010.

662   Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der

663   Werff, H., van Coillie, F. and Tiede, D.: Geographic Object-Based Image Analysis – Towards a new paradigm,

664   ISPRS J. Photogramm. Remote Sens., 87, 180–191, doi:10.1016/j.isprsjprs.2013.09.014, 2014.

665   Breiman, L.: Bagging Predictors, Mach. Learn., 24(2), 123–140, doi:10.1023/A:1018054314350, 1996.

666   Breiman, L.: Random Forests, Mach. Learn., 45(1), 5–32, doi:10.1023/A:1010933404324, 2001.

667   Curry, J. A., Schramm, J. L. and Ebert, E. E.: Sea ice-albedo climate feedback mechanism, J. Clim., 8(2), 240–247,
668   doi:10.1175/1520-0442(1995)008<0240:SIACFM>2.0.CO;2, 1995.

669   DeFries, R. .: Multiple Criteria for Evaluating Machine Learning Algorithms for Land Cover Classification from
670   Satellite Data, Remote Sens. Environ., 74(3), 503–515, doi:10.1016/S0034-4257(00)00142-5, 2000.

671   DeMott, P. J. and Hill, T. C. J.: Investigations of Spatial and Temporal Variability of Ocean and Ice Conditions in and
672   Near the Marginal Ice Zone. The "Marginal Ice Zone Observations and Processes Experiment" (MIZOPEX) Final
673   Campaign Summary, DOE ARM Climate Research Facility, Pacific Northwest National Laboratory; Richland,
674   Washington., 2016.

675   Dominguez, R.: IceBridge DMS L0 Raw Imagery, Version 1, , doi:10.5067/UMFN22VHGGMH, 2010.

676   Duro, D. C., Franklin, S. E. and Dubé, M. G.: A comparison of pixel-based and object-based image analysis with
677   selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery,
678   Remote Sens. Environ., 118, 259–272, doi:10.1016/j.rse.2011.11.020, 2012.

679   Eicken, H.: Tracer studies of pathways and rates of meltwater transport through Arctic summer sea ice, J. Geophys.
680   Res., 107(C10), 8046, doi:10.1029/2000JC000583, 2002.

681   Fetterer, F. and Untersteiner, N.: Observations of melt ponds on Arctic sea ice, J. Geophys. Res. Ocean., 103(C11),
682   24821–24835, doi:10.1029/98JC02034, 1998.

683   GDAL: GDAL - Geospatial Data Abstraction Library, Version 2.1.0, Open Source Geospatial Found. [online]
684   Available from: http://gdal.org, 2016.

685   Inoue, J., Curry, J. A. and Maslanik, J. A.: Application of Aerosondes to Melt-Pond Observations over Arctic Sea Ice,
686   J. Atmos. Ocean. Technol., 25(2), 327–334, doi:10.1175/2007JTECHA955.1, 2008.

687   Kurtz, N. T., Farrell, S. L., Studinger, M., Galin, N., Harbeck, J. P., Lindsay, R., Onana, V. D., Panzer, B. and Sonntag,
688   J. G.: Sea ice thickness, freeboard, and snow depth products from Operation IceBridge airborne data, Cryosph.,
689   7(4), 1035–1056, doi:10.5194/tc-7-1035-2013, 2013.

690   Kwok, R.: Declassified high-resolution visible imagery for Arctic sea ice investigations: An overview, Remote Sens.
691   Environ., 142, 44–56, doi:10.1016/j.rse.2013.11.015, 2014.

692   Kwok, R. and Rothrock, D. A.: Decline in Arctic sea ice thickness from submarine and ICESat records: 1958-2008,
693   Geophys. Res. Lett., 36(15), n/a-n/a, doi:10.1029/2009GL039035, 2009.

694   Landy, J., Ehn, J., Shields, M. and Barber, D.: Surface and melt pond evolution on landfast first-year sea ice in the
695   Canadian Arctic Archipelago, J. Geophys. Res. Ocean., 119(5), 3054–3075, doi:10.1002/2013JC009617, 2014.

696   Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang,
697   J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S. and Davidson, M.: CryoSat-2 estimates of Arctic
698   sea ice thickness and volume, Geophys. Res. Lett., 40(4), 732–737, doi:10.1002/grl.50193, 2013.

699   Lu, P., Li, Z., Cheng, B., Lei, R. and Zhang, R.: Sea ice surface features in Arctic summer 2008: Aerial observations,
700   Remote Sens. Environ., 114(4), 693–699, doi:10.1016/j.rse.2009.11.009, 2010.

701   Markus, T., Cavalieri, D. J., Tschudi, M. A. and Ivanoff, A.: Comparison of aerial video and Landsat 7 data over
702   ponded sea ice, Remote Sens. Environ., 86(4), 458–469, doi:10.1016/S0034-4257(03)00124-X, 2003.

703 Markus, T., Stroeve, J. C. and Miller, J.: Recent changes in Arctic sea ice melt onset, freezeup, and melt season length,
704   J. Geophys. Res., 114(C12), C12024, doi:10.1029/2009JC005436, 2009.

705 Maslanik, J., Stroeve, J., Fowler, C. and Emery, W.: Distribution and trends in Arctic sea ice age through spring 2011,
706   Geophys. Res. Lett., 38(13), doi:10.1029/2011GL047735, 2011.

707 Miao, X., Xie, H., Ackley, S. F., Perovich, D. K. and Ke, C.: Object-based detection of Arctic sea ice and melt ponds
708   using high spatial resolution aerial photographs, Cold Reg. Sci. Technol., 119, 211–222,
709   doi:10.1016/j.coldregions.2015.06.014, 2015.

710 Miao, X., Xie, H., Ackley, S. F. and Zheng, S.: Object-Based Arctic Sea Ice Ridge Detection From High-Spatial-
711   Resolution Imagery, IEEE Geosci. Remote Sens. Lett., 13(6), 787–791, doi:10.1109/LGRS.2016.2544861, 2016.

712 Pal, M.: Random forest classifier for remote sensing classification, Int. J. Remote Sens., 26(1), 217–222,
713   doi:10.1080/01431160412331269698, 2005.

714 Parkinson, C. L. and Comiso, J. C.: On the 2012 record low Arctic sea ice cover: Combined impact of preconditioning
715   and an August storm, Geophys. Res. Lett., 40(7), 1356–1361, doi:10.1002/grl.50349, 2013.

716 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss,
717   R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É.: Scikit-
718   learn: Machine Learning in Python, J. Mach. Learn. Res., 12(Oct), 2825–2830 [online] Available from:
719   http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html (Accessed 24 July 2017), 2011.

720 Perovich, D. K.: On the aggregate-scale partitioning of solar radiation in Arctic sea ice during the Surface Heat Budget
721   of the Arctic Ocean (SHEBA) field experiment, J. Geophys. Res., 110(C3), C03002, doi:10.1029/2004JC002512,
722   2005.

723 Perovich, D. K., Tucker, W. B. and Ligett, K. A.: Aerial observations of the evolution of ice surface conditions during
724   summer, J. Geophys. Res., 107(C10), 8048, doi:10.1029/2000JC000449, 2002a.

725 Perovich, D. K., Grenfell, T. C., Light, B. and Hobbs, P. V: Seasonal evolution of the albedo of multiyear Arctic sea
726   ice, J. Geophys. Res., 107(C10), 8044, doi:10.1029/2000JC000438, 2002b.

727 Perovich, D. K., Richter-Menge, J. A., Jones, K. F. and Light, B.: Sunlight, water, and ice: Extreme Arctic sea ice
728   melt during the summer of 2007, Geophys. Res. Lett., 35(11), L11501, doi:10.1029/2008GL034007, 2008.

729 Petty, A. A., Tsamados, M. C., Kurtz, N. T., Farrell, S. L., Newman, T., Harbeck, J. P., Feltham, D. L. and Richter-
730   Menge, J. A.: Characterizing Arctic sea ice topography using high-resolution IceBridge data, Cryosph., 10(3),
731   1161–1179, doi:10.5194/tc-10-1161-2016, 2016.

732 Pistone, K., Eisenman, I. and Ramanathan, V.: Observational determination of albedo decrease caused by vanishing
733   Arctic sea ice, Proc. Natl. Acad. Sci., 111(9), 3322–3326, doi:10.1073/pnas.1318201111, 2014.

734 Polashenski, C., Perovich, D. and Courville, Z.: The mechanisms of sea ice melt pond formation and evolution, J.
735   Geophys. Res. Ocean., 117(C1), n/a-n/a, doi:10.1029/2011JC007231, 2012.

736 Renner, A. H. H., Gerland, S., Haas, C., Spreen, G., Beckers, J. F., Hansen, E., Nicolaus, M. and Goodwin, H.:
737   Evidence of Arctic sea ice thinning from direct observations, Geophys. Res. Lett., 41(14), 5029–5036,
738   doi:10.1002/2014GL060369, 2014.

739 Rösel, A. and Kaleschke, L.: Comparison of different retrieval techniques for melt ponds on Arctic sea ice from
740   Landsat and MODIS satellite data, Ann. Glaciol., 52(57), 185–191, doi:10.3189/172756411795931606, 2011.

741 Rösel, A., Kaleschke, L. and Birnbaum, G.: Melt ponds on Arctic sea ice determined from MODIS satellite data using
742   an artificial neural network, Cryosph., 6(2), 431–446, doi:10.5194/tc-6-431-2012, 2012.

743 Stroeve, J. C., Serreze, M. C., Holland, M. M., Kay, J. E., Malanik, J. and Barrett, A. P.: The Arctic's rapidly shrinking
744   sea ice cover: a research synthesis, Clim. Change, 110, 1005–1027, doi:10.1007/s10584-011-0101-1, 2012.

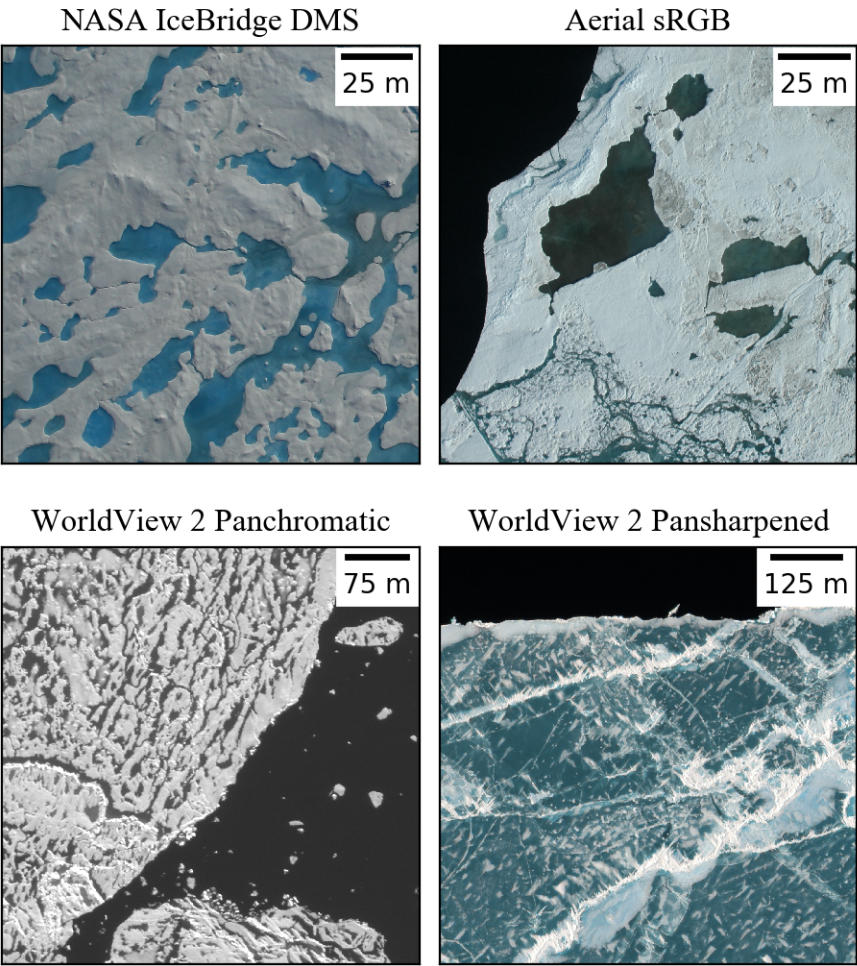
745 Stroeve, J. C., Markus, T., Boisvert, L., Miller, J. and Barrett, A.: Changes in Arctic melt season and implifications
746   for sea ice loss, Geophys. Res. Lett., 41, 1216–1225, doi:10.1002/2013GL058951.Received, 2014.

747 Tschudi, M. A., Maslanik, J. A. and Perovich, D. K.: Derivation of melt pond coverage on Arctic sea ice using MODIS
748   observations, Remote Sens. Environ., 112(5), 2605–2614, doi:10.1016/j.rse.2007.12.009, 2008.

749 van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E. and Yu,
750   T.: scikit-image: image processing in Python, PeerJ, 2, e453, doi:10.7717/peerj.453, 2014.

751 Webster, M. A., Rigor, I. G., Perovich, D. K., Richter-menge, J. A., Polashenski, C. M. and Light, B.: Seasonal
752   evolution of melt ponds on Arctic sea ice, J. Geophys. Res. Ocean., 120(9), 1–15,
753   doi:10.1002/2015JC011030.Received, 2015.

754 Yan, G., Mas, J. -F., Maathuis, B. H. P., Xiangmin, Z. and Van Dijk, P. M.: Comparison of pixel-based and object-
755   oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia, China, Int. J.
756   Remote Sens., 27(18), 4039–4055, doi:10.1080/01431160600702632, 2006.

757

  Figures

759

760 **Figure 1. Examples of imagery from each of the four imaging platforms that we seek to classify in this study. Each type of**
761 **imagery has either a different spatial resolution or and different levels spectral information available.**

762

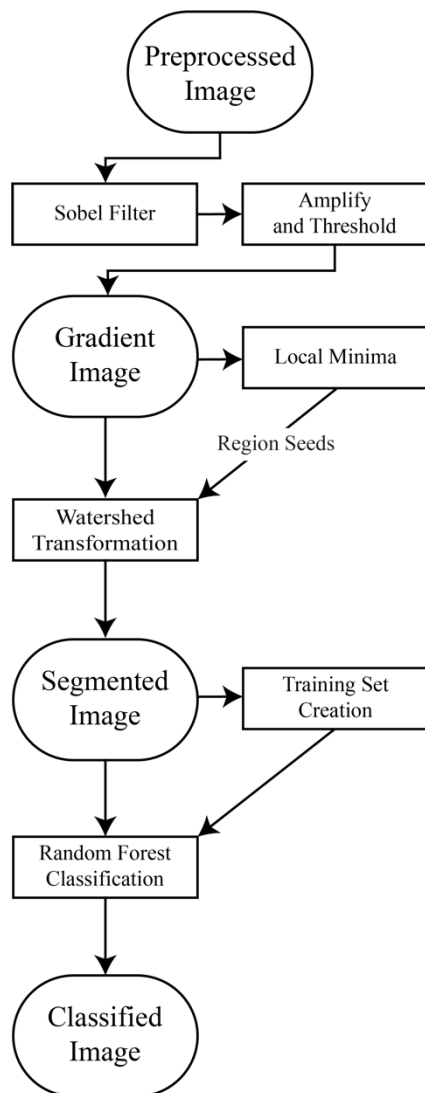



Figure 2. Flow diagram depicting the steps taken to classify an image in the OSSP algorithm.

764

765

Raw Image   Image Segments   Classification Output

☐ Snow + Ice   ◼ Dark and Thin Ice   ◼ Melt Ponds + Submerged Ice   ◼ Open Water

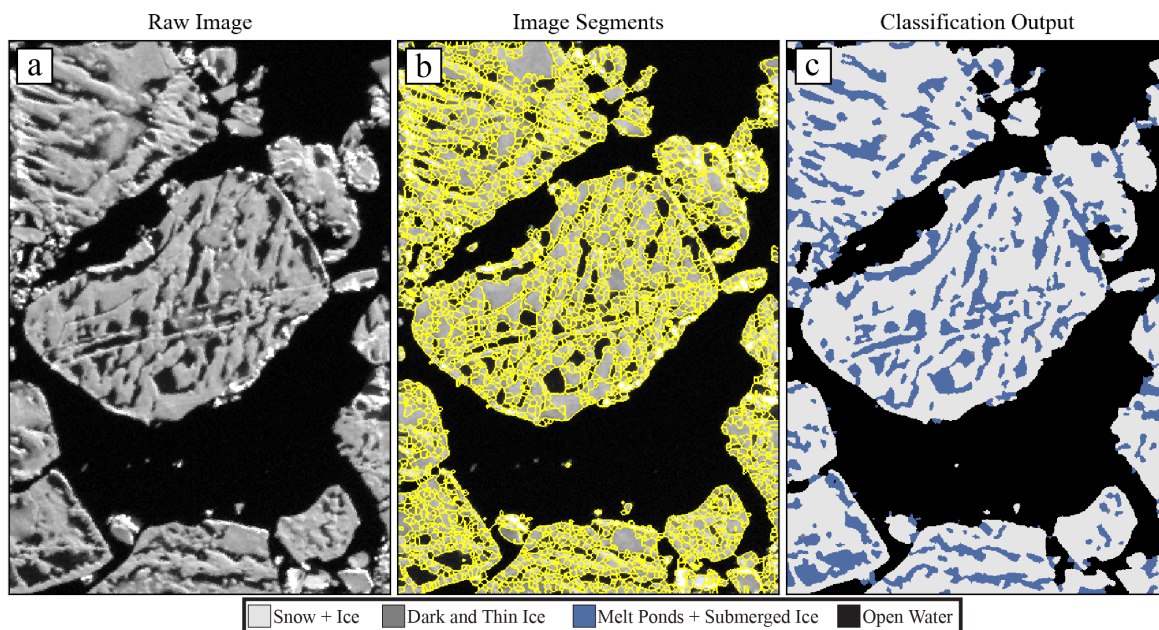



Figure 3. Visual representation of important steps in the image processing workflow. Panel (a) shows preprocessed panchromatic WorldView 2 satellite imagery, taken on July 1, 2014. In panel (b), outlines of the image objects created by our edge detection and watershed transformation are shown overlain on top of the image in panel (a). Panel (c) shows the result of replacing each object with a value corresponding to the prediction of the random forest classifier.

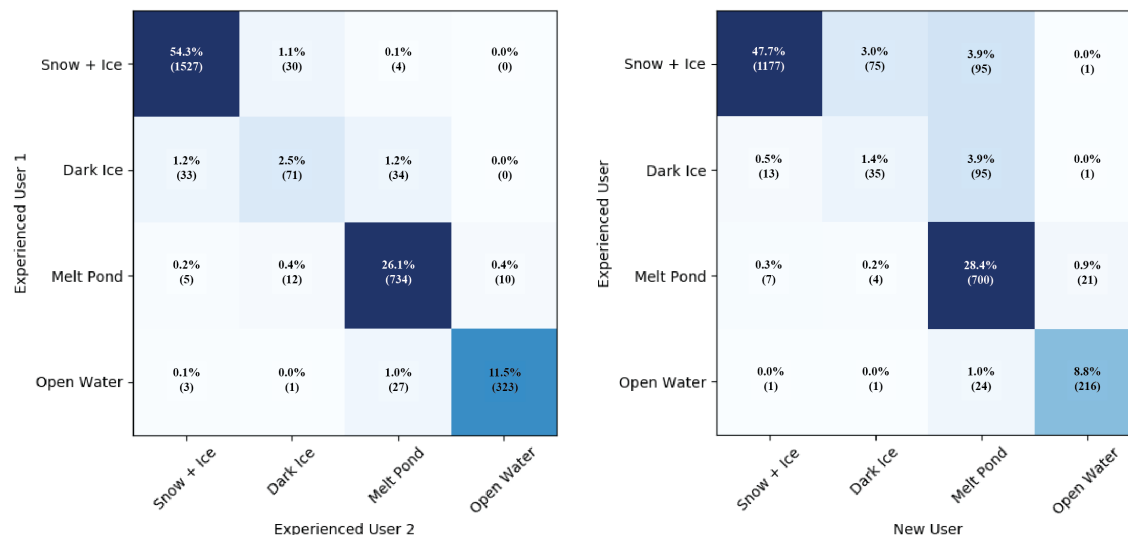



Figure 4. Confusion matrices comparing classification tendencies between two users experienced with the image processing algorithm (left) and between an experienced user and a new user (right). Squares are colored based on the value of the cell, with darker colors indicating more matches. Values along the diagonal of each confusion matrix represents the agreement between each user, while values in off-diagonal regions represent disagreement.

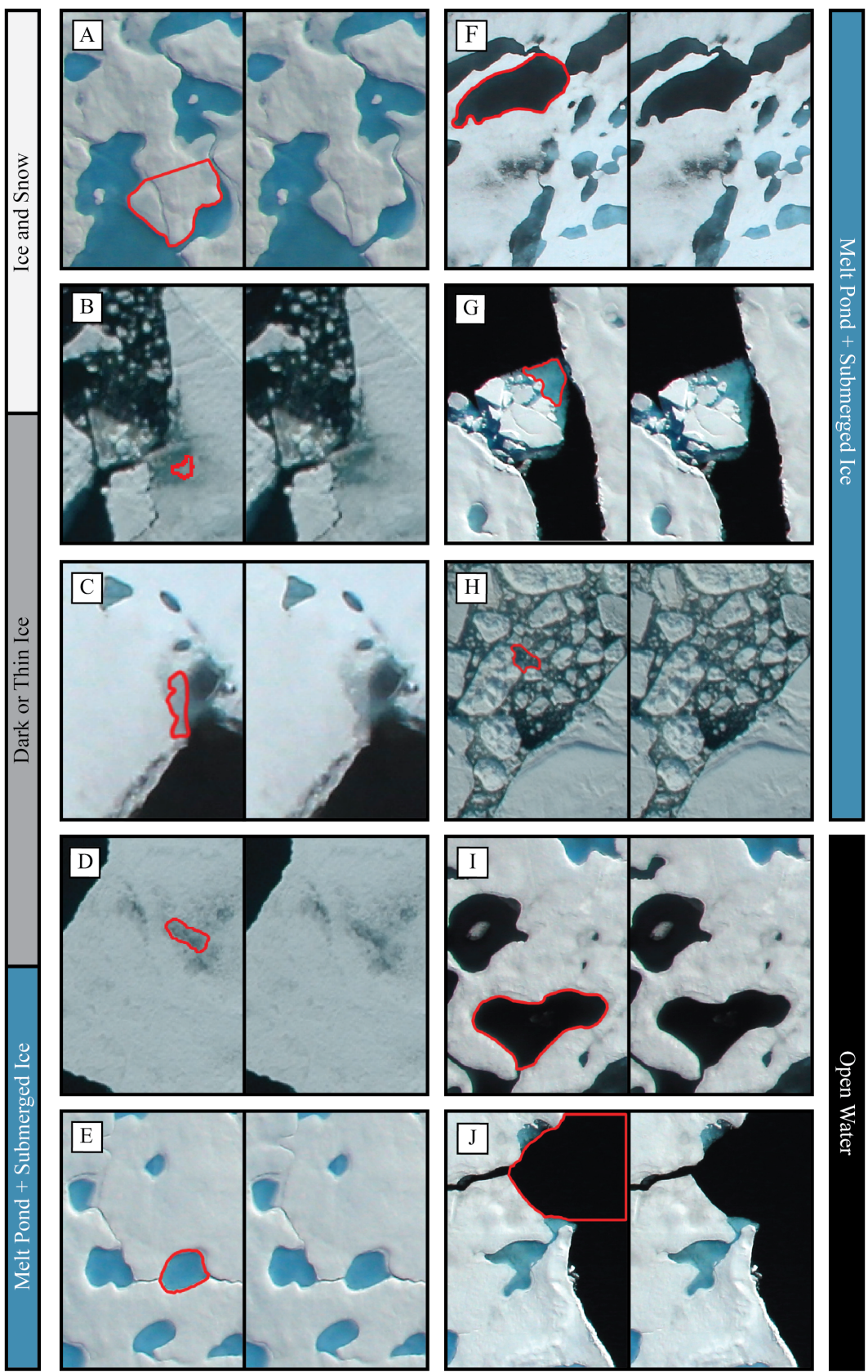



Figure 5. Examples of surfaces seen in aerial imagery of sea ice that span our four classification categories. Panel A: snow covered surface. Panel B: Ice with a thin surface scattering layer where disagreement on true classification exists - represents a small fraction of total surface area. Panel C: Panel D: Surface transitioning to a melt pond that is not yet fully submerged. Panel E: Melt pond. Panel F: Dark melt pond that has not completely melted through. Panel G: Submerged ice. Panel H: Brash, mostly submerged, included in the melt pond category. Panel I: Melt pond that has completely melted through to open water. Panel J: Open water.

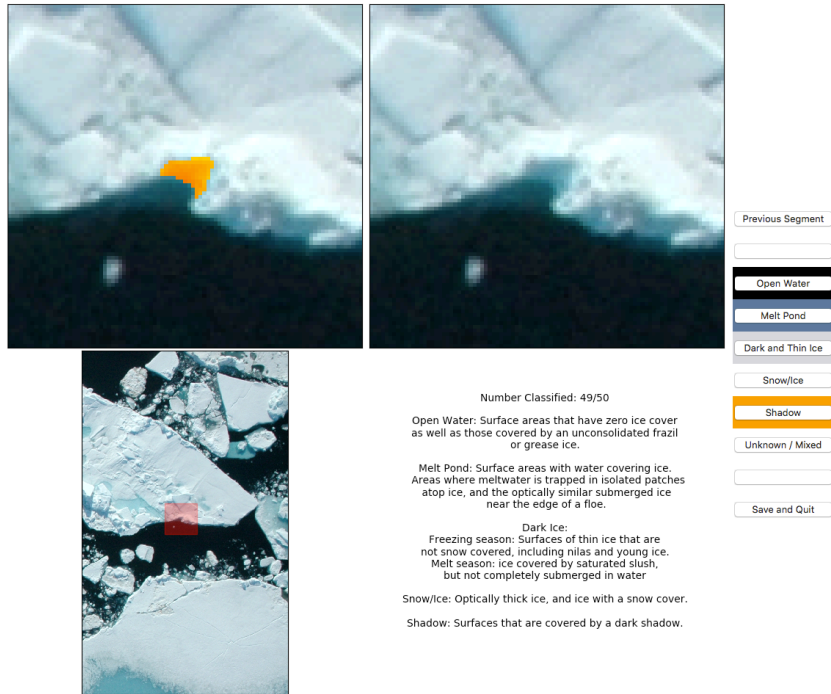
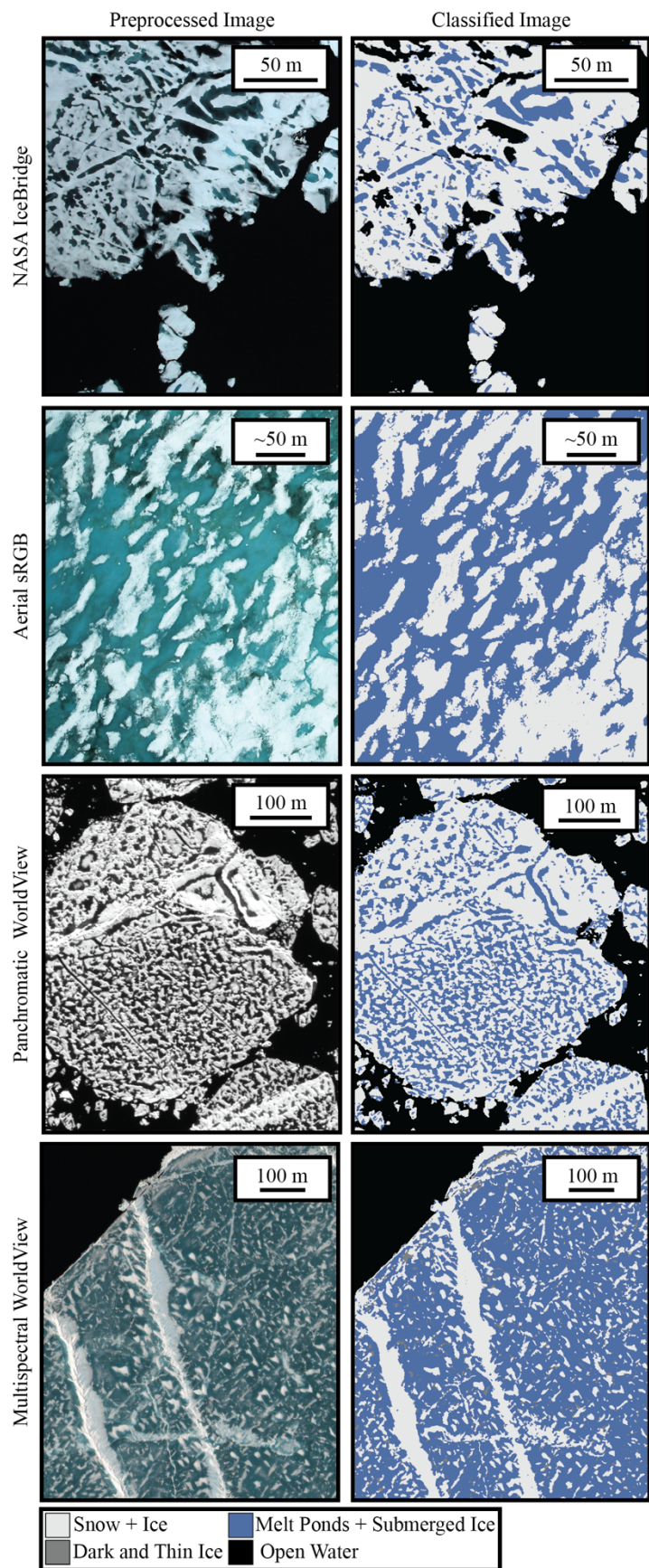
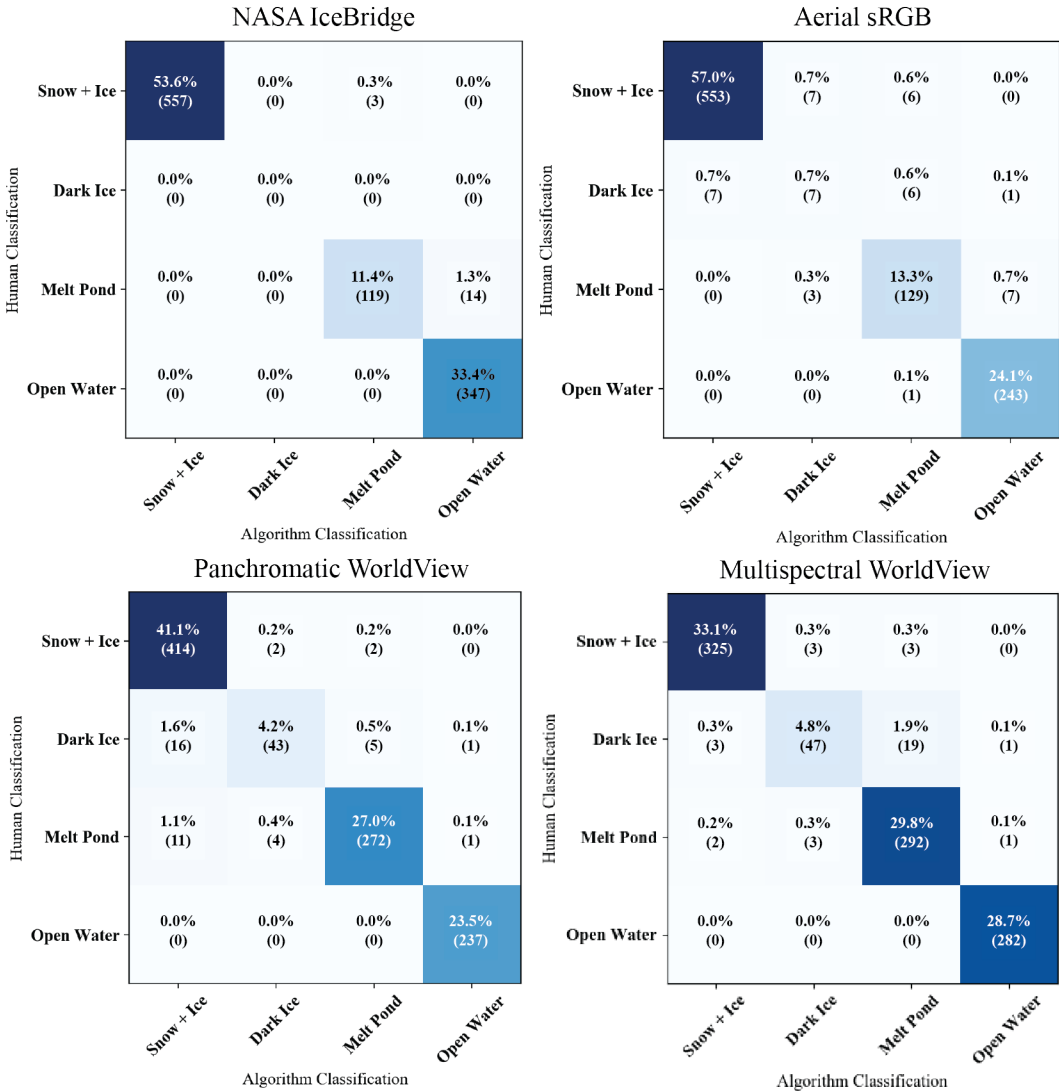
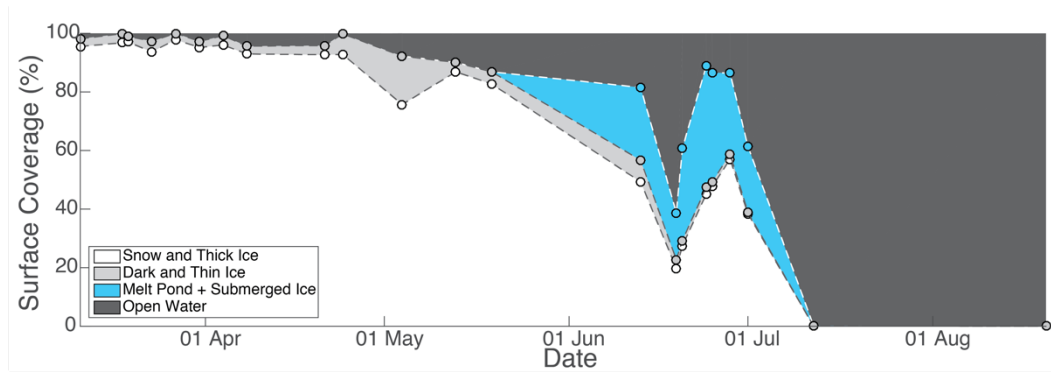
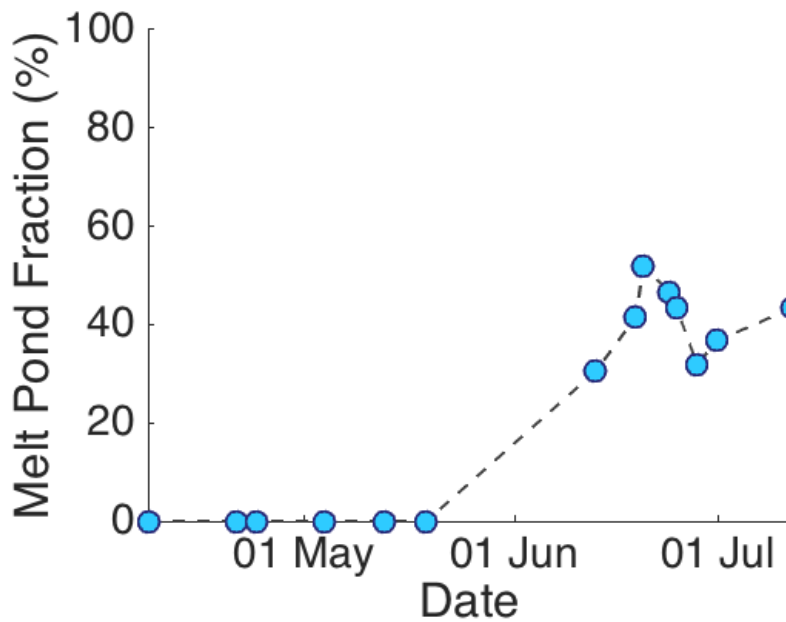



Figure 6. Graphical user interface used to create training datasets and to assess the accuracy of a classified image. Bottom left panel shows an overview of the region to provide the user with spatial context. Top left magnifies the image and highlights the segment of interest, while top right shows the same region with no segment overlap. The user is allowed to choose between any of the relevant surface categories, or to indicate that they are unsure of the classification. As shown, the user interface is demonstrating the classification of a segment for use in a training set. This same GUI is also capable of asking a user to classify an individual pixel, which can be compared to the final classified image for determining accuracy (section 3.6).

793 **Figure 7. Side-by-side comparison of preprocessed imagery (left) and the result of classification (right) for each of the four**
794 **imaging platforms. Images depict ice surfaces in varying stages of melt. The NASA IceBridge image, for example, is in very**
795 **late stages of melt ponds that have already melted through to the ocean.**

796

797 **Figure 8. Accuracy confusion matricies comparing the classification of 1000-pixels between a human and the algorithm.**
798 **Squares are colored based on the value of the cell, with darker colors indicating more matches. Values along the diagonal**
799 **of each confusion matrix represents the agreement between each classifier, while values in off-diagonal regions represent**
800 **disagreement.**

801

802 **Figure 9. Seasonal progression of surface type distributions at the satellite image collection site; 2014 in the Beaufort Sea**
803 **at 72°N 128°W. This site represents a Eulerian observation of the sea ice surface, and does not track a floe across its lifetime.**
804 **Average scene size was 956km^2 with a minimum of 304km^2 and a maximum of 1321km^2 .**

805

806

807 **Figure 10. Evolution of melt pond fraction over the 2014 season at our satellite image collection site; 2014 in the Beaufort**
808 **Sea at 72°N 128°W. This site represents a Eulerian observation of the sea ice surface, and does not track a floe across its**
809 **lifetime. By August, the sea ice extent has retreated north of this location, and we therefore do not capture a full melt pond**
810 **cycle.**

811

812

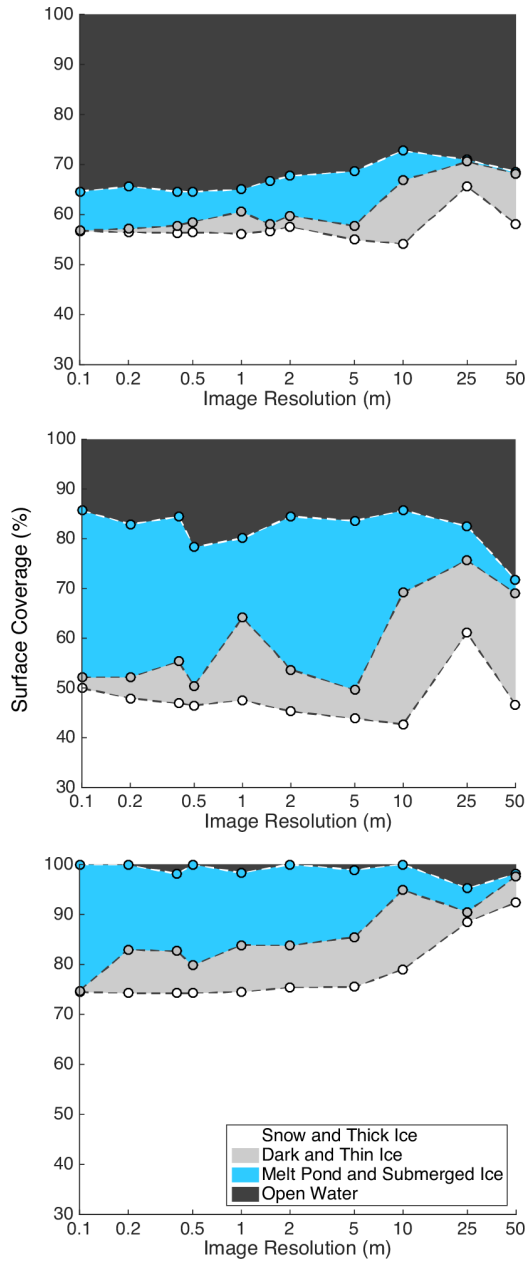
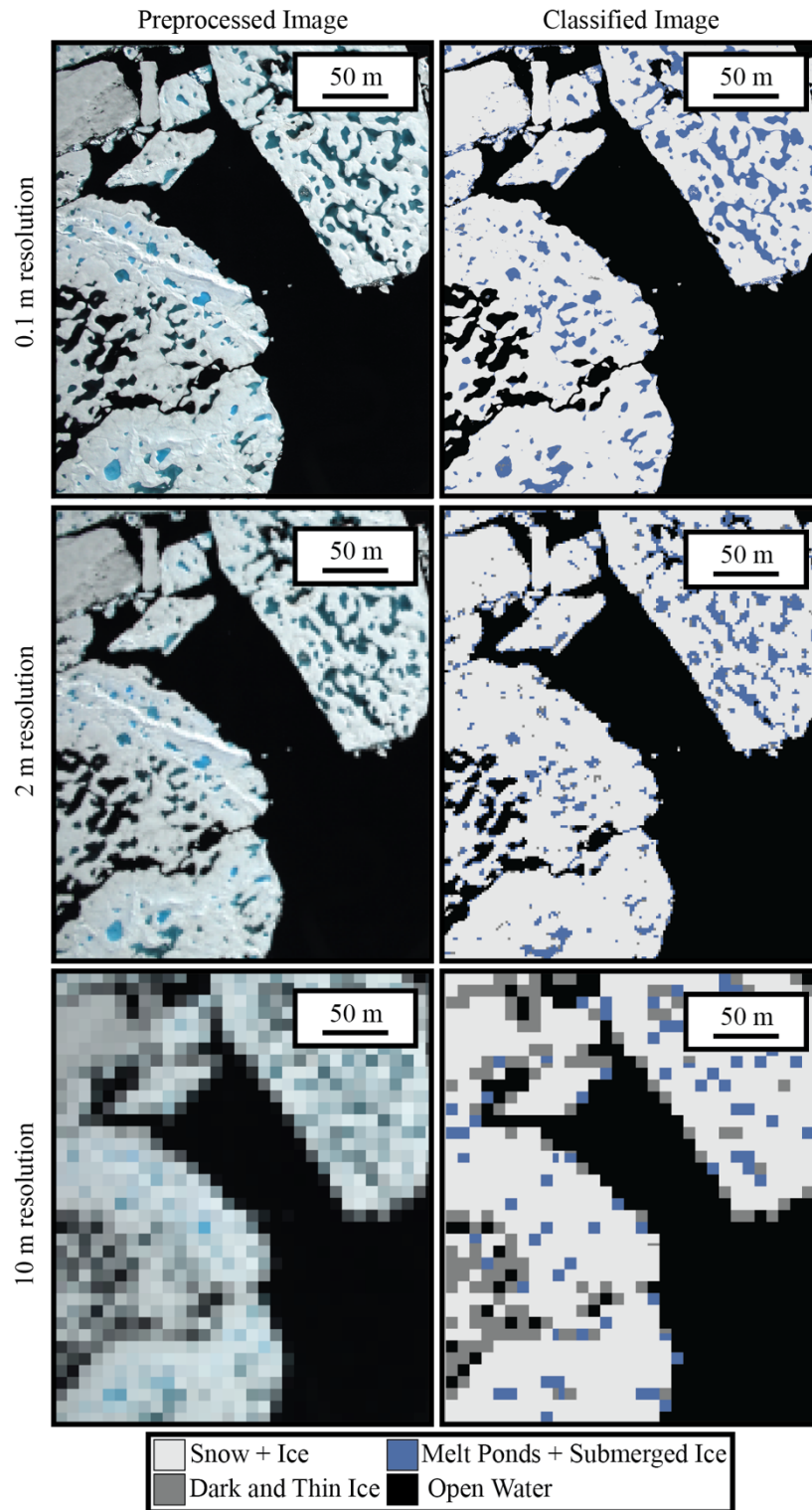



Figure 11. Change in surface coverage percentage as a result of downsampling three IceBridge images. Each plot represents
813
814 **a single image, with resolution along the x-axis on a log scale. Imagery starts at the nominal IceBridge resolution of 0.1m**
815 **and is degraded to a maximum of 50m.**

816
817 **Figure 12. Visual demonstration of the downsampling effect on a single NASA IceBridge image. The top image is shown at**
818 **the original 0.1 m resolution. The middle image is a resolution of 2m - the equivalent of a multispectral WorldView 2 image**
819 **without pansharpening. The bottom has a resolution of 10m, where pixel size has begun to exceed the average melt pond**
820 **size.**

821

822

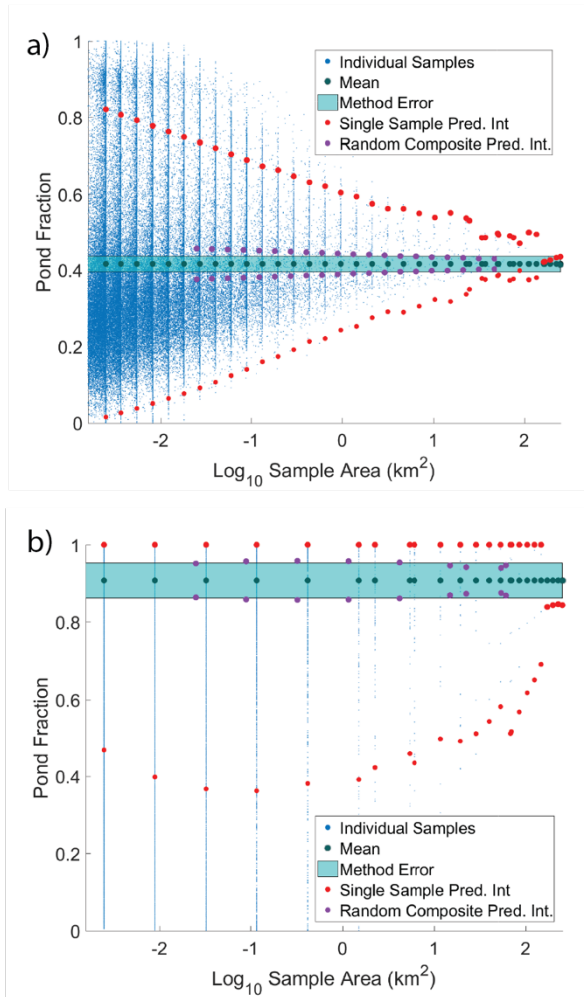
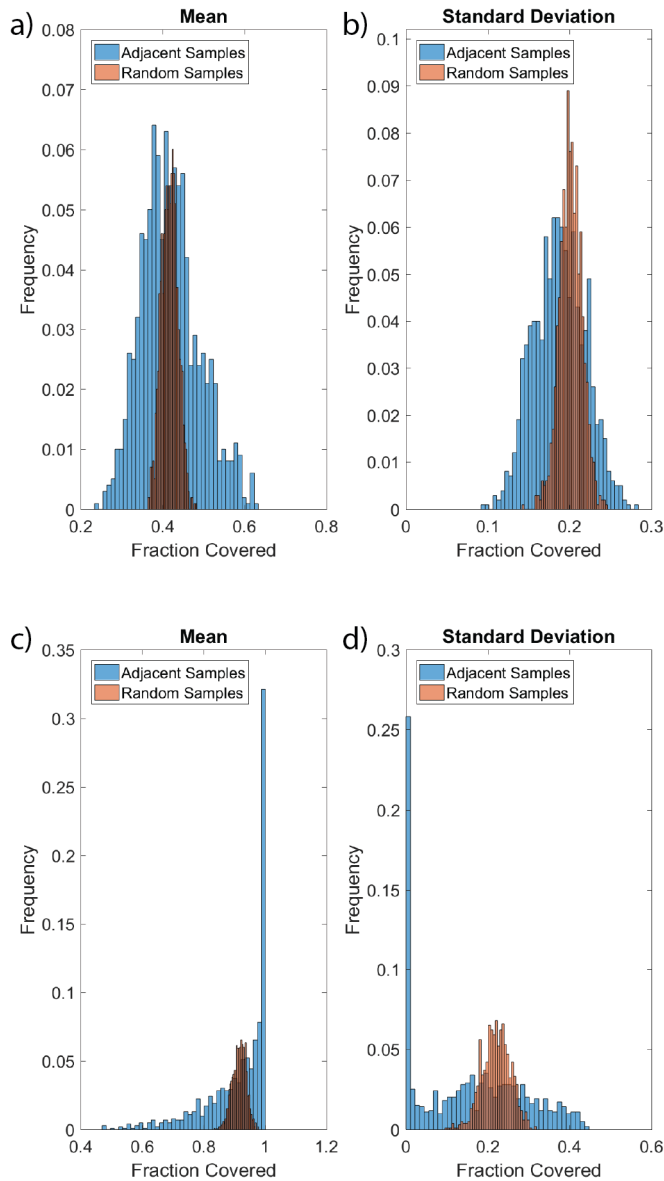



Figure 13. Convergence of melt pond fraction (a) and ice fraction (b) for a WorldView image collected 25 June 2014 at 72°N 128°W as the area evaluated is increased. Small blue dots represent individual image subsets. For segments of a given size, black dots represent the mean value of those samples, red dots represent the 95% prediction interval, and purple dots show the 95% prediction interval for the same total area, but calculated from 100 randomly placed, smaller, samples. Cyan shaded area represents the error in determination expected from the processing method.

828

829 **Figure 14. Histogram of mean (a) and standard deviation (b) of 1000 melt pond fraction estimates, each calculated from**
830 **100 sample areas on a 25 June 2014 WorldView image. The 100 samples were either randomly distributed across the image**
831 **(red) or adjacent to each other (blue). Panels (c) and (d) show the same as (a) and (b), respectively, for ice fraction rather**
832 **than melt pond fraction.**

833

834

835

836

837

838

Tables

Attribute	MS	PAN	Aerial
Mean (Pan)	Light gray	Blue	Light gray
Mean (Coastal)	Blue	Light gray	Light gray
Mean (Blue)	Blue	Light gray	Blue
Mean (Green)	Blue	Light gray	Blue
Mean (Yellow)	Blue	Light gray	Light gray
Mean (Red)	Blue	Light gray	Blue
Mean (Red Edge)	Blue	Light gray	Light gray
Mean (NIR1)	Blue	Light gray	Light gray
Mean (NIR2)	Dark gray	Light gray	Light gray
Median (Pan)	Light gray	Blue	Light gray
StDev (Pan)	Light gray	Blue	Light gray
Min Intensity (Pan)	Dark gray	Blue	Dark gray
Max Intensity (Pan)	Dark gray	Blue	Dark gray
StDev (Blue)	Dark gray	Light gray	Blue
StDev (Green)	Dark gray	Light gray	Blue
StDev (Red)	Dark gray	Light gray	Blue
Entropy	Dark gray	Blue	Blue
Segment Size	Dark gray	Blue	Blue
Image Date	Blue	Blue	Blue
Coastal / Green	Blue	Light gray	Dark gray
Blue / NIR1	Blue	Light gray	Light gray
Green / NIR1	Blue	Light gray	Light gray
Yellow / Red Edge	Blue	Light gray	Light gray
Yellow / NIR1	Blue	Light gray	Light gray
Yellow / NIR2	Blue	Light gray	Light gray
Red / NIR1	Blue	Light gray	Light gray
(B1 - NIR1)/(B2 + NIR1)	Blue	Light gray	Light gray
(G - R)/(G + R)	Blue	Light gray	Light gray
(B - R)/(B + R)[1]	Dark gray	Light gray	Blue
(B - G)/(B + G)[1]	Dark gray	Light gray	Blue
(G - R)/(2*B - G - R)[1]	Dark gray	Light gray	Blue
Neighbor Mean	Blue	Blue	Blue
Neighbor StDev	Dark gray	Blue	Blue
Neighbor Max	Dark gray	Blue	Blue
Neighbor Entropy	Dark gray	Blue	Blue

[1]**Miao et al. 2015**

Table 1. Attributes used for classifying each of the three image types. Blue squares indicate attributes that were used for that image. Dark gray squares indicate attributes that are available, but were not found to be sufficiently beneficial in the classification to merit inclusion under our criteria. Light gray squares indicate attribute that are not available on that image type (e.g. band ratios on a panchromatic image). NIR is the near infrared wavelength. B1 is the costal WorldView band, and B2 is the blue band. R, G, and B, stand for red, green, and blue, respectively.

Image ID	Sensor Type	Date Collected	Ice + Snow	DTI	MPS	OW	Accuracy
102001002C214D00	Panchromatic	11-Mar-14	96	3	0	2	97
103001002E8F0D00	Panchromatic	18-Mar-14	97	3	0	0	97
102001002BBA0C00	Panchromatic	19-Mar-14	97	2	0	1	96
103001002FC75200	Panchromatic	23-Mar-14	94	4	0	3	95
102001002CB77C00	Panchromatic	27-Mar-14	98	2	0	0	100
1030010030403A00	Panchromatic	31-Mar-14	95	2	0	3	98
1030010031B65000	Panchromatic	4-Apr-14	96	3	0	1	99
102001002BA6C100	Panchromatic	8-Apr-14	93	3	0	4	100
103001002F79A700	Panchromatic	21-Apr-14	93	3	0	4	98
1030010030371B00	Panchromatic	24-Apr-14	93	7	0	0	98
103001003102A600	Panchromatic	4-May-14	76	16	0	8	98
102001003007FA00	Panchromatic	13-May-14	87	3	0	10	97
10300100306F2E00	Panchromatic	19-May-14	83	4	0	13	96
102001003035D700	Panchromatic	13-Jun-14	49	7	25	18	95
1030010033AAC400	Panchromatic	19-Jun-14	20	3	16	61	97
1020010031DF9E00	Panchromatic	20-Jun-14	27	2	31	39	96
1020010032B94E00	Panchromatic	24-Jun-14	45	2	41	11	95
102001003122A700	Panchromatic	25-Jun-14	48	1	37	13	97
102001002F4F1A00	Panchromatic	28-Jun-14	57	2	28	14	95
10300100346D1200	Panchromatic	1-Jul-14	38	0	23	39	97
1030010035C8D000	Panchromatic	12-Jul-14	0	0	0	100	100
103001003421AB00	Panchromatic	20-Aug-14	0	0	0	100	100
10300100324B7D00	Multispectral	13-Jun-14	44	7	29	19	96
1030010033AAC400	Multispectral	19-Jun-14	16	3	19	62	97
10300100346D1200	Multispectral	1-Jul-14	44	2	26	28	98
1030010035C8D000	Multispectral	12-Jul-14	0	0	0	100	100
2016_07_13_05863	IceBridge	13-Jul-16	50	2	34	14	92
2016_07_13_05882	IceBridge	13-Jul-16	72	1	26	0	97
2016_07_13_05996	IceBridge	13-Jul-16	70	2	28	0	95
2016_07_13_06018	IceBridge	13-Jul-16	61	2	36	1	91
2016_07_13_06087	IceBridge	13-Jul-16	66	1	33	0	99
2016_07_16_00373	IceBridge	16-Jul-16	9	0	2	89	100
2016_07_16_00385	IceBridge	16-Jul-16	66	1	14	20	98

2016_07_16_00662	IceBridge	16-Jul-16	49	1	16	35	98
2016_07_16_00739	IceBridge	16-Jul-16	67	2	25	6	97
2016_07_16_01569	IceBridge	16-Jul-16	22	0	7	71	97
2016_07_16_02654	IceBridge	16-Jul-16	35	0	10	54	95
2016_07_19_01172	IceBridge	19-Jul-16	62	0	14	24	90
2016_07_19_01179	IceBridge	19-Jul-16	57	0	10	32	95
2016_07_19_02599	IceBridge	19-Jul-16	51	0	7	43	99
2016_07_19_02603	IceBridge	19-Jul-16	69	0	9	22	99
2016_07_19_02735	IceBridge	19-Jul-16	74	0	25	0	100
2016_07_19_03299	IceBridge	19-Jul-16	57	0	8	35	96
2016_07_21_01221	IceBridge	21-Jul-16	49	0	4	47	97
2016_07_21_01311	IceBridge	21-Jul-16	87	1	5	7	95
2016_07_21_01316	IceBridge	21-Jul-16	92	0	4	4	99
DSC_0154	Aerial sRGB	8-Jun-09	43	4	53	0	94
DSC_0327	Aerial sRGB	8-Jun-09	33	3	63	0	90
DSC_0375	Aerial sRGB	8-Jun-09	96	0	4	0	99
DSC_0422	Aerial sRGB	8-Jun-09	88	0	11	0	98
DSC_0223	Aerial sRGB	10-Jun-09	46	1	53	0	93
DSC_0243	Aerial sRGB	10-Jun-09	59	1	40	1	98
DSC_0314	Aerial sRGB	10-Jun-09	89	0	11	0	95
DSC_0319	Aerial sRGB	10-Jun-09	75	2	19	4	88
DSC_0323	Aerial sRGB	10-Jun-09	37	2	61	0	95
DSC_0338	Aerial sRGB	10-Jun-09	83	2	15	1	95
DSC_0386	Aerial sRGB	10-Jun-09	80	3	14	3	89
DSC_0394	Aerial sRGB	10-Jun-09	79	2	10	9	95
DSC_0412	Aerial sRGB	10-Jun-09	63	2	24	10	92
DSC_0425	Aerial sRGB	10-Jun-09	56	2	17	24	97
DSC_0439	Aerial sRGB	10-Jun-09	71	1	6	22	98
DSC_0441	Aerial sRGB	10-Jun-09	57	0	4	38	98
DSC_0486	Aerial sRGB	10-Jun-09	53	1	17	29	96
DSC_0634	Aerial sRGB	10-Jun-09	72	1	14	12	96
DSC_0207	Aerial sRGB	13-Jun-09	80	1	19	0	96
DSC_0514	Aerial sRGB	13-Jun-09	86	1	13	0	97

847 **Results Table 2. The complete results of imagery processed for this analysis. Descriptions for each image includes the image**
848 **type, date collected, the percent of the image that falls into each of the four categories, and the accuracy assessment.**

849

Image Source	Training Dataset Size	Out-of-bag Error
Panchromatic WorldView	1000	0.94
Pansharpened WorldView	859	0.89
Aerial Imagery	945	0.94
IceBridge Imagery	940	0.91

850 **Table 3. Out-of-Bag scores for the three training datasets used to classify imagery from each of the four sensor platforms,**
851 **and the number of objects manually classified for each set.**