Impact of Spatial, Spectral, and Radiometric Properties of Multispectral Imagers on Glacier Surface Classification

Allen POPE[a,*,1], W. Gareth REES[a]

[a] *Scott Polar Research Institute, University of Cambridge, Lensfield Road, Cambridge CB2 1ER, United Kingdom*

**Corresponding author, allen.pope@post.harvard.edu, +1 (303) 492-9077*

Keywords: glaciers, snow, multispectral, classification, principal component analysis.

Abstract

Using multispectral remote sensing, glacier surfaces can be classified into a range of zones. The properties of these classes are used for a range of glaciological applications including mass balance measurements, glacial hydrology, and melt modelling. However, it is not immediately evident that multispectral data should be optimal for imaging glaciers and ice caps. Thus, this investigation takes an inverse perspective. Taking into account spectral and radiometric properties, *in situ* spectral reflectance data were used to simulate glacier surface response for a suite of multispectral sensors. Sensor-simulated data were classified and compared. In addition, airborne multispectral imagery was classified for a range of spatial resolutions and intercompared in three different ways. In these analyses, the most important property which determined the suitability of a multispectral imager for glacier surface classification was its radiometric range (i.e. gain settings). Low resolution imagery (250 m pixels) is too coarse to represent the true complexity present on a glacier while medium resolution imagery (60 m, 30 m, or 20 m)

[1] *Present Affiliations: National Snow and Ice Data Center, University of Colorado, Boulder, Boulder, CO, USA & Department of Earth Sciences, Dartmouth College, Hanover, NH, USA*

accurately represented the results derived from high resolution airborne imagery. Of those studied here, the satellite imagers currently in use that are most suitable for glacier surface classification are Landsat TM / ETM+ and ASTER (all with particular gain settings). Both Sentinel-2 and the OLI on Landsat 8 are also expected to be similarly qualified. Landsat MSS is also found to be radiometrically well-suited for glacier surface classification, but its lower spatial resolution makes it a secondary selection.

1. Introduction

The world's glaciers and ice caps (GIC), which respond much more quickly to shifting climate than the continental ice sheets, provide information about past and present climate variability, are central parts of the world's hydrological cycle, and are key to understanding regional and global climate change (e.g. Cogley et al., 2011; Oerlemans, 1994). In addition, glaciers contribute to local biodiversity (Jacobsen et al., 2012), mediate the hydrology and flooding of some mountain systems (Dahlke et al., 2012), and provide crucial water resources for large populations of the world (Baraer et al., 2011; Barry, 2011; Björnsson & Pálsson, 2008; Bolch et al., 2012; Hopkinson & Demuth, 2006).

Glacier surface properties are integral to the behaviour of GIC. The division of GIC into accumulation and ablation areas is just the beginning of classification of glacier facies, or zones (Benson, 1960; Williams et al., 1991). The equilibrium line altitude (ELA) and accumulation area ratio (AAR; Cogley et al., 2011) can be used as proxies for glacier mass balance (Braithwaite, 1984; Dyurgerov, 1996). In addition, the glacier surface controls much of a glacier's energy balance (Cuffey & Patterson, 2010). Energy balance models both assimilate remotely sensed data about glacier surfaces to improve their results (Machguth et al., 2009; van Angelen et al., 2012), as well as validate their results (Braun et al., 2007; de Woul et al., 2006).

Multispectral imagery is often the best tool for studying glaciers surfaces (Pellikka & Rees, 2010). Reflectance information over a range of wavelengths, good spatial resolution, frequent repeat imaging, an extensive image archive, and often cost-free data access are all important. However, multispectral sensors were not designed by glaciologists. Satellites like the original Landsat were (and continue to be) designed for a range of tasks including agricultural, oceanographic, and atmospheric monitoring (Markham & Helder, 2012). Therefore, it is not self-evident that they should be optimal for imaging GIC. Thus, the roles that the various spectral, spatial, and radiometric properties of each sensor play in the success and output of resulting classifications remain unquantified.

1.1 Research Aims

Multispectral imagers are powerful tools, and with the increasing availability of a range of high quality multispectral data including the recently launched Landsat 8 (Irons et al., 2012) and upcoming Sentinel 2 (Drusch et al., 2012), it is increasingly crucial that they are fully understood. This investigation therefore takes an inverse perspective; it aims to start with *in situ* data to investigate the extent of information available from full-spectrum data and what that means for efficient and consistent application across multispectral sensors with different band capabilities and combinations. We ask the questions: How do the spectral and radiometric properties of these sensors limit or enhance their performance in glacier classification? Because sensors are characterised by both spatial and spectral properties, how does the spatial resolution of these various sensors impact the resultant surface classification? And what does this mean for glaciological applications?

2. Background

2.1 Glacier Facies

Glacier surfaces exhibit a range of zones – wet and dry, snowy and icy, clean and dirty. In order to understand better the changing conditions of a glacier's surface, the area can be considered to be divided into a set of systematic, idealised facies that are characterised by a particular set of properties relating to the metamorphosis of the snow or ice surface; facies range from dry snow at colder, higher elevations through to melting ice near the glacier terminus (Benson, 1960; Williams et al., 1991). Although there is a wide range of glacier facies, a glacier can be divided into two larger regions: the accumulation zone and the ablation zone. The transition between these two areas is the line of net zero annual mass change known as the equilibrium line (Cogley et al., 2011; Cuffey & Patterson, 2010). Each different configuration of facies is evidence of a different metamorphic history. Facies distributions vary across glaciers both within seasons and across years, and not all facies are necessarily present on all glaciers.

In addition, beyond these zones which are considered 'facies,' further surface classes can be identified *in situ* and remotely. For example, there are extensive areas of wind glaze and sastrugi in Antarctica (Kuchiki et al., 2011; Orheim & Lucchitta, 1987; Scambos et al., 2012). The presence of snow algae imparts a reddish tinge to an evolving wet-snow facies (Takeuchi, 2009), and dust or black carbon will darken the snow surface (e.g. Painter, 2011). Debris cover on the glacier can be considered another type of surface class (e.g. Casey et al., 2012; Shukla et al., 2009), as can volcanic ash deposited on glacier surfaces from a nearby eruption. In this study, 'facies' are considered to be the idealised zones of glacier surfaces which relate directly to accumulation and melt, while 'surface classes' are the zones which can be distinguished from the surface.

Identification of accumulation versus ablation classes (through the ELA or the AAR) can be used as a proxy for a glacier's mass balance, often in combination with further data such as a digital elevation model (e.g. Braithwaite & Müller, 1978; Dyurgerov, Meier, & Bahr, 2009; Rabatel et al., 2005; Shea et al., 2013). Also, glacier facies can be related to mass balance in other ways. Snow is bright (i.e. highly reflective in much of the visible and near-infrared) and ice is darker, therefore as the melt season progresses the glacier as a whole gets darker overall - specifically in proportion to the relative contributions of different glacier facies. In this way, it is possible to monitor glacier albedo as a tool for monitoring glacier mass balance (Dumont et al., 2012; Greuell & Oerlemans, 2005; Greuell et al., 2007).

Shortwave radiation is crucial to the energy balance of a glacier. Glacier facies meaningfully contribute to this radiation balance and therefore to the surface energy balance of GIC. A clear example of the interrelated nature of energy balance and glacier facies can be seen in the simple parameterization of the degree-day melt model where ice and snow have different degree day factors (e.g. Hock, 2003). Information about the interannual and intra-annual evolution of glacier surfaces is also a key parameter in building energy balance models. Fuller consideration of glacier facies in glacier melt modelling is gaining increasing traction within the glaciological community (e.g. Dumont et al., 2010; Machguth et al., 2009). This is true not just for GIC, but also for the larger ice sheets, where better classification and description of the unique properties of different facies improve melt model behaviour (e.g. van Angelen et al., 2012).

Snow and ice reflectances are heavily wavelength-dependent (e.g. Wiscombe & Warren, 1980). In particular, the NIR (near infrared, ~700-1400 nm) has been seen as containing quantitative information about snow and ice surfaces (Kokhanovsky & Zege, 2004; Li et al., 2001; Nolin & Dozier, 1993). Glacier facies classification, too, has focused on the NIR to the exclusion of the visible, although snow studies have highlighted both ranges (Zeng et al., 1983). Sidjack and Wheate (1999) and Braun et al. (2007) cited some saturation in the visible and enhanced performance in the NIR as reasons for choosing linear combinations of input multispectral bands which contained large contributions from the NIR and SWIR (shortwave infrared, ~1400-2500 nm) and minimal contributions from the visible.

Based on these examples, it is natural to hypothesise that sensors with enhanced capabilities in the NIR will be able to classify glacier facies better than their counterparts. This belief will be investigated below.

2.2 Multispectral Remote Sensing, Classification, and Glacier Facies

Multispectral remote sensing images are some of the most prevalent, easily available, and versatile forms of data available for the Earth Observing glaciologist. There are a variety of factors which must be weighed in choosing an appropriate multispectral sensor; each separate investigation or task requires an imager which is fit for purpose. Major considerations include spatial resolution, spectral resolution (i.e. band wavelengths), radiometric resolution and range, temporal resolution (i.e. revisit time), data cost and ease of access, length of data archive, data availability, and availability of pre-processed products. From the range of different options, it is highly unlikely that any one sensor will be optimal for all studies. Nevertheless, sensors were chosen to span a wide range of properties (i.e. spatial scales, spectral bands, and gain settings) and priority was placed on wide use and easy data access.

Although many imagers could have been included, those not included (e.g. SPOT, WorldView, etc.) will be able to find analogous properties in those considered here. Figure 1 includes the range of popular and prominent multispectral imagers that are considered in this study.

INSERT Figure 1 approximately here

Glaciological uses of multispectral imagery include glacier maps and inventories (Albert, 2002; Hendriks & Pellikka, 2008; Kargel et al., 2005; Paul & Kääb, 2005; Paul, 2000), albedo calculation (Greuell et al., 2002; Knap, Reijmer, & Oerlemans, 1999), distinguishing snow from cloud (Hall et al., 1995), identification of surface and basal crevasses (Luckman et al., 2012), feature tracking (Heid & Kääb, 2012), interpolating digital elevation models (Pope et al., 2013), identifying ice sheet grounding lines (Bindschadler et al., 2011), and much more (Pellikka & Rees, 2010; Rees, 2006).

Classification, the process that takes quantitative information from every pixel and places each into one of a group of discrete categories, is crucial for image interpretation. Many different techniques have been applied to multispectral data to identify glacier surface classes. It should be noted that (automated) classification of glacier extent is considered to be a separate problem, one which has been largely solved, with the exception of debris-covered areas (Paul et al., 2013). Unsupervised classifications have had significant success in classifying glacier facies not only because they are easily reproducible but also because they are often able to exploit subtle features within data sets. ISODATA (Iterative Self-Organizing Data Analysis; e.g. Aniya et al., 1996; De Angelis et al., 2007; Nolin & Payne, 2007; Sidjak & Wheate, 1999; Wolken et al., 2009) and k-means classification (e.g. Barcaza et al., 2009; König et al., 2004) are the most widely and easily implemented clustering algorithms for glacier surface classification. This study will therefore implement these two iterative, unsupervised classification techniques.

2.2.1 Impact of Spatial Resolution on Multispectral Classification

Data is not bad or good because of coarse or fine resolution, but difference scale imagery is better or worse suited to particular applications. Coarse imagery often confounds land cover classification as a result of pixels representing a mixture of classes. In this case, increasing resolution would increase classification success. Resolution effects have been investigated in a range of different applications (Baker et al., 2013; Battersby et al., 2012; Michishita et al., 2012; Phinn et al., 2012; Sobrino et al., 2012). The results of these comparisons have shown that agreement between spatial scales varies depending on the type of surface and the scale of inhomogeneity. In other words, although it is not immediately intuitive, increasing spatial resolution can decrease classification success. This happens in the case where coarser imagery serves to smooth out spatial inhomogeneity within classes.

To the authors' knowledge, no study has directly considered this subject for glaciers. Previous studies have used higher resolution imagery (~10 and ~1 m) to assess the accuracy of glacier extent measurements using medium resolution (~30 m) imagery (Paul, 2000; Paul et al., 2013). There is no significant difference in measured glacier area using imagery at 60 m resolution and finer; lower resolution imagery was not tested (Paul et al., 2002). In glacier albedo calculations, the scale of albedo variations is smaller than 30 m Landsat pixels (Reijmer et al., 1999), but because albedo can vary within facies this does not necessarily mean glacier facies vary on the same scales. Albedo variations within facies would then confound glacier surface classifications with spatial resolution finer than 30 m.

3. Field Sites and Data

3.1 Field Spectra

For this study, visible through shortwave infrared (350-2500 nm) hemispherical-directional reflectances (HDRs; Schaepman-Strub et al., 2006) were collected during two field campaigns. In August 2010, data were collected on Midtre Lovénbreen, Svalbard. In August 2011, data were collected on the major western outlet of Langjökull, Iceland. These two locations were chosen so as to include sampling of glaciers that had undergone different accumulation and melting histories. Additional differences were introduced by the springtime eruption of Grímsvötn volcano in Iceland. The data, as well as consideration of available snow and ice spectral reflectance measurements and modelling efforts, are fully presented in Pope and Rees (In Press) and are also available upon request to the corresponding author.

3.2 Multispectral Imagery

Imagery collected with the UK Natural Environment Research Council (NERC) Airborne Research and Survey Facility's (ARSF) Airborne Thematic Mapper (ATM) was used to investigate the effects of spatial resolution. The ATM was designed to mimic many of Landsat 7 ETM+'s bands but with added spectral coverage and spatial resolution (30 vs. ~2 m, determined by flight height and processing; see Figure 1).

Insert Figure 2 approximately here

The ARSF flew a campaign over Midtre Lovénbreen on 9 August 2003 (see Figure 2); the ATM was mounted inside the ARSF's Dornier 228 aircraft. Simultaneously collected laser ranging data (Rees & Arnold, 2007) were used to orthorectify the imagery. Azgcorr version 5.0.0, produced by Azimuth Systems UK and provided by the ARSF, was used to perform the orthorectification (Azimuth Systems, 2005). ATM measurements are delivered as at-sensor radiance. No measurements of incoming radiation were available coincident with airborne data collection, and calibration to reflectance with pseudo-invariant off-glacier targets from Landsat imagery was attempted, but no nearby surface was found to be consistent in its reflectance. Therefore, ATM data were left as at-sensor radiance. No surface anisotropy or slope corrections were implemented. The meteorological records at the nearby Ny-Ålesund research station also indicate that the glacier surface froze overnight and that frost deposition was likely. Angular crystals on the surface may increase surface reflectance (Casacchia et al., 2001). These combined effects may have some impact on glacier surface classification. However, as this study is concerned mainly with intercomparison of classifications at different resolutions, and the ATM image is only compared to spatially degraded versions of itself, possible classification-confounding factors will be self-consistent and therefore should not impact the results of this study.

4. Methods

4.1 Spectral Response Matching

A narrow to broadband (NTB) conversion is necessary to use spectral reflectance to replicate multispectral reflectance. This is done with a known relative spectral response function for each band obtained via NASA, ESA, and the NERC Field Spectroscopy Facility. Sentinel-2 and OLI are both approximated with "top hat functions" (i.e. uniform spectral response across each band), as no further data were available. In addition, spectra are not used to simulate Landsat-8 OLI data because calibration values were not available at the time of writing. The 12 bit radiometric resolution of the OLI is assessed by comparing the effect of radiometric resolution on the full-spectrum data. In addition, only MODIS bands 1-16 are considered because higher bands are in unsuitable wavelengths (e.g. water absorption) and their spectral response functions were not available. More detail on spectral response matching is available in the supplementary material.

4.2 Multispectral Sensor Radiometric Properties

Beyond the NTB conversion, further calibration parameters from each multispectral band must be taken into account to fully simulate the measurements which would be taken by a multispectral sensor – had the user been holding an extremely portable version of a satellite rather than a FieldSpec's photodiodes. Full details on how field data were used to simulate sensor radiometric properties are available in the supplementary material.

4.3 Principal Component Analysis

The spectra measured in this study are highly multidimensional data, and multispectral data have a number of dimensions (in this case, bands) themselves. That is to say, each data "point" is characterised by many values rather than a single number. Therefore, some method is needed to reduce the dimensionality of the data for analysis – without losing important information – in order to understand what the most important input wavelengths are for glacier surface classification. Concisely put, PCA is a transformation that reduces the dimensionality of a dataset by reprojecting it into a new coordinate space (e.g. Boresjö Bronge & Bronge, 1999; Sidjak & Wheate, 1999). Thus, each principal component (PC) is a linear combination of the input data.

The first two or three PCs produced from full-spectrum reflectance can be used to create transferrable linear combinations (LCs) which are optimised for particular satellite bands for glacier surface using appropriate relative response functions (Pope & Rees, In Press). Here, PCs were calculated separately for Langjökull and Midtre Lovénbreen field spectra, and coefficients were rounded and compared. This has two benefits: one, it aids in a conceptual understanding of what each LC is emphasizing within the data; two, it facilitates wider transferability of LCs by not being specifically tailored to the field data. A unified set of LCs was produced for each satellite for later analysis (see below). For glaciers, LC1 is representative of VNIR albedo, LC2 emphasizes the difference in reflectance at blue / green wavelengths and red / NIR wavelengths, and LC3 highlights the difference in blue / NIR reflectance and green / red reflectance. All LCs are available in the supplementary material. LC1 in each case is representative of VNIR albedo, LC2 emphasizes the difference in reflectance at blue / green wavelengths and red / NIR wavelengths, and LC3 highlights the difference in blue / NIR reflectance and green / red reflectance. All LCs are available in the supplementary material.

4.4 Classification

Both ATM data and LCs are clustered using very similar techniques. Following previous studies (Braun et al., 2007; de Woul et al., 2006; Pope & Rees, In Press), an arbitrary number of classes were identified with a clustering algorithm, ISODATA for ATM data and k-means for LCs, respectively. For ATM imagery, output classes were subsequently grouped into accumulation and ablation areas for statistical analysis. For LCs, grouped classes are discussed below.

4.5 Statistical Classification Comparison

This study requires that two main classification accuracy assessments be conducted. The first is a clustering analysis where the classification of field spectra is compared to knowledge from fieldwork. The second compares the results of classification of ATM imagery degraded to different spatial resolutions. A contingency matrix is created which shows the number of times (i.e. number of pixels or number of spectra) where two classifications agree or disagree. This matrix provides information on errors of both omission and commission and is the basis for statistical analysis (Congalton & Green, 1999; Congalton, 1991; Foody, 2002; Rees, 2008).

From the information in the contingency matrix, the most basic statistic is "*A*," or the overall accuracy agreement. This is the sum of the times for which the classifications agree divided by the total number of samples. Put another way, *A* is the trace of the normalised contingency matrix (Rees, 2008). However, random chance can lead to agreement of classes, and therefore *A* can overestimate classification accuracy.

In response, Cohen's Kappa (*K*, Cohen, 1960) is a statistic which accounts for random effects within the classification comparison and remains an indexed value (i.e. perfect agreement result in $K = 1$). In this way, *K* reduces the overestimation of classification success included in *A*. *K* values can also be described by qualitative descriptions rather than simply numerical values (Monserud & Leemans, 1992).

5. Results & Interpretation

This study aims to answer the question of what qualities define the best multispectral imagers for glacier surface classification. Therefore, this section begins with spectral and radiometric considerations, transitions to an investigation of the impact of spatial resolution, and then combines the two to understand the advantages and limitations of a range of popular multispectral sensors to glacier surface classification.

5.1 Impact of Spectral and Radiometric Properties on Glacier Surface Classification

5.1.1 Midtre Lovénbreen Clustering and Classification Analysis

For this study, Midtre Lovénbreen provides an example of a largely "clean" and simple glacier. LCs 1, 2, and 3 are calculated for each sensor; all cases are described in Table 1 and presented in Figure 3. The data were merged into three larger groups defined as snow surfaces, ice surfaces, and wet surfaces; these three main classes are circled in Figure 3a. Classification success is assessed, and the results presented in Table 1 are ranked in order of descending *K*.

Insert Table 1 and Figure 3 approximately here

Overall, all sensors in all settings are largely able to classify the three clusters. Even the worst case attains almost 80% success and is considered "very good." This success is tempered by the fact that the classification task is idealistically easy because there are no pixels as there would be in real imagery. There is one setup which does not group the wet classes as well as the other sensors: MODIS using bands 8 and higher in a low gain setting (Figure 3o). This is possibly due to lack of contributions to the LCs from NIR wavelengths.

Radiometric resolution on its own does not appear to be important in the "clean glacier" case on Midtre Lovénbreen, as can be seen by comparing Figures 1a, 1d, 1e, and 1c which are all clustered using full spectra but at unrestricted, 16 bit, 12 bit, and 8 bit resolution, respectively. However, when combined with a limited set of bands to use in LCs, the quantisation does begin to appear (e.g. Equations 11 and 12 in the supplementary material used to produce Figures 1g and 1i, emulating ASTER). This example supports the importance of using LCs with a higher number of band combinations which better represent the full spectrum surface reflectance.

Radiometer properties (i.e. radiance range and gain settings) do appear to play some role in classification accuracy, but it is hard to assess fully with such widespread success for the Midtre Lovénbreen spectra. Saturation is clearly visible in some Landsat settings; its distinctive signature is a linear alignment of spectra with high LC1 values for Landsat TM/ ETM+ HHHH (i.e. high gain for Bands

1-4) and Landsat MSS in Figures 1l and 1f, respectively. Landsat TM/ETM+ LLLH replicates original spectra biplots better than Landsat LLLL (i.e. low gain for Bands 1-4; Figures 1k and 1p, respectively), which makes sense given the high to lower reflectance transition of glacier surfaces from visible to NIR wavelengths. Interestingly, ASTER in a high gain setting (Figure 3b) appears to have such success in classification because all of the bright, snowy surfaces were compressed into the same saturated point in the plot, thereby making the entire class almost entirely homogeneous.

Ultimately, while varying slightly in performance, imager radiometric properties do not provide significant limitations or guidance in selecting the most appropriate multispectral imager for surface classification of clean glaciers.

5.1.2 Langjökull Clustering and Classification Analysis

Langjökull provides an example of a glacier with a more complex set and larger range of surface classes, furnished in this case by ash from the Grímsvötn eruption in spring 2011. For Langjökull spectra, a very similar analysis to Midtre Lovénbreen is performed. All cases are described in Table 2 and presented in Figure 4. However, the results were merged into only two classes (clean ice and other; this is indicated by the circle in Figure 4a). Again, classification is assessed using knowledge from fieldwork, and the results are presented in Table 2 ranked in order of descending *K*.

Insert Table 2 and Figure 4 approximately here

While the “clean” glacier ice was easily classified across all sensors, and despite the apparently simpler task of dividing into two groups rather than Midtre Lovénbreen’s three, there is a much larger range of success between sensors and settings for the Langjökull spectra. For the ash-covered glacier, no imager emulation is fully able to represent the range of information contained in the full spectrum data. From the spread of data points, it appears this is largely due to lack of sensitivity in LC2, although saturation in LC1 also plays a role. Nevertheless, because of the simpler task (i.e. identifying clean ice), it is possible to achieve "perfect" classification success (for terminology see Monserud & Leemans, 1992).

Radiometric resolution, again, is not found to be important on its own, but the quantising effects are again seen in ASTER because of the smaller number of bands contributing to the LCs (see Figures 2g, 2i, and to a lesser extent 2n). Even without restricting spectral range and rounding for radiometric resolution only (no scaling, 16 bit, 12 bit, and 8 bit in Figures 2a, 2b, 2c, and 2d, respectively), saturation does appear to be present in the data; this is because some spectra were measured at higher than 100% reflectance as a result of a specular component of reflectance (see Section 3.5.2).

Indeed, radiometric resolution is overshadowed by other factors. For example, 12 bit MODIS sensors would be expected to perform well given their higher radiometric resolution compared to Landsat's 8 bits. However, MODIS gain settings are tuned for darker land and ocean surfaces and so are less suitable to the task of glacier surface classification. As this demonstrates, radiometric range is more important than radiometric resolution.

The importance of radiometric range and gain settings is also demonstrated by ASTER and Landsat. ASTER hi gain (Figure 4n), Landsat TM / ETM+ LLLL (Figure 4m), Landsat MSS (Figure 4h), and Landsat TM / ETM+ HHHH (Figure 4k) all show a linear feature influencing the higher ranges of both LC1 and LC2, the result of "sensor" saturation. This is more pronounced for ASTER than Landsat because more contributing "bands" are saturated, and LC2 shares more coefficients in common with LC1 for ASTER. Landsat MSS is actually very similar to Landsat TM / ETM+ HHHH, but Landsat LLLH performs better than other Landsat setups. For ASTER, as in the Midtre Lovénbreen case, the brightest classes (New Drifted Snow 1, New Drifted Snow 3, and White Ice 2) are compressed to a single point.

The LC1-LC2 biplots for MODIS 8+ in both gain settings (Figure 4o and p) demonstrate an intriguing chevron shape. Various theories were considered for why this would occur, including lack of band representation in the NIR or perhaps particular placement of bands in higher and lower wavelengths causing anomalous effects in LC2. In order to find the real culprit, it is necessary to return to exactly what LC1 and LC2 are (see Equations 26 and 17 in supplementary material). Unlike for other bands, the magnitude of all coefficients of bands contributing to the LCs for MODIS 8+ are identical, except for a flip in sign for higher bands. Where reflectance in Bands 8 to 12 is much higher than 13 to 16, there is a positive linear pattern, and when the opposite is true there is a negative linear pattern. The linear pattern is more pronounced than it would be otherwise because snow and ice spectra have fairly uniform reflectance across the range of wavelengths observed by Bands 8 to 16. Although PCs are uncorrelated, rounding in the coefficients of the LCs caused an artefact in this case.

Ultimately, as can be seen with the *K* rankings in Table 2, while a large number of sensors do a "perfect" job identifying clean ice on Langjökull, others perform very poorly. There is a pronounced division between the two, jumping from $K = 0.9081$ for ASTER normal gain down to 0.6579 for MODIS 1-7. It should be noted that these results hold only for a simple unsupervised classification; supervised or iterative approaches have the potential to yield more specific classes but would lose transferability and ease of implementation. For surface classification of "dirty" or ash-covered glaciers, these results indicate that sensors with "Good," "Fair," or "Poor" *K* values should be foregone in preference for the many alternative sensors which rank higher in performance.

5.2 Impact of Spatial Resolution on Glacier Surface Classification

5.2.1 Experimental Strategy and Considerations

ATM imagery of most of Midtre Lovénbreen was used in this experiment. LCs were calculated for the 2 m imagery (Equations 8 and 9 in supplementary material), and the image is masked using a manual outline of the glacier. LCs were degraded to 20, 30, 60 and 250 m pixels analogous to Sentinel-2, Landsat TM / ETM+ / OLI, Landsat MSS, and MODIS Bands 1-2, respectively; 500 m imagery is considered too coarse to resolve smaller mountain glaciers. LCs were then input into an ISODATA classification (10 classes, maximum 10 iterations, 95% convergence); output classes were merged by the user into meaningful glaciological classes (i.e. ablation and accumulation facies).

As spatial resolution is varied, the radiometric content of all images remains constant as a control. As alluded to earlier, the impact of resolution will depend on the fractal scale of the surfaces being considered. For this purpose, Midtre Lovénbreen is taken to be a representative glacier surface because there is no reason to believe otherwise. Most classification accuracy and quality assessments suffer from their inability to fulfil some assumptions, namely co-registration and random sampling (e.g. Comber, Fisher, Brunsdon, & Khmag, 2012); this experiment uses all pixels to assess accuracy, and because coarser imagery is created from finer imagery, co-registration is a not an issue.

By definition, the detail available in the 2 m results will be blurred out (to the lower resolution), but it is unknown what beneficial or detrimental effects this may have on surface classification accuracy and what this will mean for individual sensors which provide imagery at a variety of spatial resolutions. In this case, only pixel-based classification is considered because pixel-based classifications have been traditionally easier to implement with a range of software tools. Although object-based image analysis has been shown to have benefits for very-high resolution imagery (1 m), at any lower resolutions it does not produce statistically significantly different results from pixel-based classification (Baker et al., 2013).

5.2.2 Spatial Resolution Assessment 1

For each resolution, the ISODATA classification outputs 10 classes (see Figure 5). Classes 1 and 2 are mixes of shadow and thin debris cover, classes 3 through 8 are interpreted as ablation classes, and classes 9 and 10 are interpreted as accumulation classes. For reasons described in earlier sections concerning the potential presence of frost and the use of radiance rather than reflectance, from investigation of the visible imagery it appears this interpretation (i.e. merging of classes) may slightly overestimate the accumulation area. Nevertheless, this is deemed to be preferable to significant underestimation of the area of accumulation facies.

Percentages of each class and the aggregated accumulation and ablation areas are presented in Table 3. Most classifications have very similar accumulation and ablation area measurements, with the exception of the 250 m pixel classification, which results in slightly less ablation and more accumulation, although the differences are below 5%. While these figures agree, that does not mean the classification results agree on the pixel level; to understand this, further analysis is necessary.

Insert Table 3 and Figure 5 approximately here

5.2.3 Spatial Resolution Assessment 2

To begin to understand pixel-level agreement of glacier surface classification at various resolutions, a majority filter was used to down-sample high resolution images to lower resolutions (see Table 4). Because class numbers are not indicative mathematically of their similarity or difference, mathematic convolution would not be meaningful. As would be expected, similarity is the most meaningful between images with similarly sized pixels (see Table 4). However, the 2 and 30 m images and 2 and 20 m images are more similar than the 20 and 30 m images. While this could be the result of a resampling artefact, it does indicate that medium resolution imagery is doing a good job at reproducing the results obtained with high resolution imagery. The 2 m classification results are approximately as similar to the 60 m results as all of the medium resolution classifications are to each other, potentially indicating that 20 or 30 m imagery is more suited to glacier surface classification than 60 m resolution imagery. At the bottom, the low resolution imagery (250 m pixels) by a large step shows the lowest agreement with all other results; according to this result, MODIS imagery is not appropriate for glacier surface classification.

The *A* and *K* rankings of the classification comparisons differ slightly in the relative position of the comparison of 60 m and higher resolution results, 250 m to 2 m and 20 m images. The inferences drawn above are consistently supported by both *A* and *K*, but divisions are more visible in the *K* values than in the *A* values. This lends some confidence to the conclusions, because *K* values should contain more signal and less false agreement than *A* values.

Insert Table 4 approximately here

5.2.4 Spatial Resolution Assessment 3

In addition to downscaling high resolution imagery, the low resolution imagery is resampled to high resolution by converting each pixel in the high resolution image to a host of small pixels of the same class (i.e. one 250 m pixel becomes 15,625 corresponding 2 m pixels). Pixels in the original high resolution and the down-sampled classifications are compared and assessed using *A* and *K* (see Table 5). *A* and *K* ranks and relative values are in better agreement than in the previous section. For this round of comparisons, similarity in pixel size was the unambiguous driver of similarity in results. Again, there is a clear break between high (2 m) / medium (20, 30, or 60 m) resolution image results and any comparison to the 250 m resolution classification results.

Insert Table 5 approximately here

6. Discussion

Using full-spectrum *in situ* reflectance data to emulate the spectral and radiometric properties of a range of imagers and settings is a controlled experiment of sorts, removing uncertainty introduced by unknown or changing surface conditions. For both clean and dirty glacier surfaces, although radiometric resolution is largely insignificant, selecting the sensor / gain settings with the most appropriate radiometric range is very important. For the data presented here, Landsat TM / ETM+ LLLH, Sentinel-2, Landsat MSS, ASTER low 1, and ASTER normal perform the classification tests the best. The radiometric properties of the recently-launched OLI were not available at the time of writing, but by considering analogues for its spectral bands and radiometric resolution, it is possible to envision that it would yield results which are a cross between full-spectrum 12 bit results, Sentinel-2, and Landsat TM / ETM+ and would therefore be quite well suited to glacier surface classification, too.

Moving on to spatial resolution, each method used to compare glacier surface classification at different pixel sizes gives a slightly different impression of the importance of sensor spatial resolution. Figure 5 clearly shows the loss of detail associated with observation at lower resolutions, but the relative area of shadow, debris, accumulation and ablation facies is very similar among images of all classes.

However, relative accuracy at different spatial scales is dependent on the scale of inhomogeneity within and between classes. For Midtre Lovénbreen, classes appear to be similarly behaved at high and medium resolutions, but the glacier is definitely more complex than 250 m pixels can capture. Although 15 m pixels (analogous to ASTER bands or fused ETM+ images) are not explicitly considered, in view of these results, such an analysis would appear to have been superfluous. Based upon this analysis, for glacier surface classification, high resolution imagery would indeed be desirable. It appears that even the highest resolution that MODIS is capable of providing (250 m) is insufficient for glacier surface classification. Medium resolution imagery is found to be adequate for the task, and 20 or 30 m imagery is preferable to 60 m imagery but not drastically.

However, it is important to question how representative the surfaces of Langjökull (spectrally) and Midtre Lovénbreen (both spectrally and spatially) are of glaciers in general. The selection of field spectra sampling locations was based upon the exploration of the field party. For Midtre Lovénbreen, the glacier is small and therefore it is highly unlikely that any major classes were omitted. Langjökull is much larger, and measurements were limited to a single outlet. Nevertheless, Landsat classification of this outlet indicates the presence of the full range of facies along the transect which was used, and therefore it is unlikely that any major classes were omitted there, either.

The question then turns to the relative proportion of each class as measured; to an extent, it is important to recognize that these relative proportions will have some impact on the statistics, in particular the proportions in any contingency table. The ranking of sensors according to K values could conceivably have been more impacted by different proportions of facies. For example, for Midtre Lovénbreen, inclusion of more 'coarse snow' and 'dry ice' spectra would likely have depressed all K values. Similarly, for Langjökull, including more spectra from the classes near the 'white ice' spectra could have had a similar impact. However, although the magnitude of the statistics would have changed, this would have impacted (beneficially or detrimentally) all simulated clustering analyses, and it is therefore unlikely that the conclusions thereof (based on relative rank) would change. This study based conclusions on all available data, choosing not to filter out spectra. Nevertheless, potential impacts of relative proportions of classes would be testable with further sampling campaigns.

It is crucial for this study that the spectra measured on Midtre Lovénbreen and Langjökull (and principal components thereof) are considered to be representative of other glaciers. Because spectra are a result of a combination of physical processes (snow formation, accumulation, compaction, melt, etc.), it is reasonable to expect that this will be the case. Indeed, the similarity of the first and second principal components of spectra between the two glaciers supports this interpretation. In addition, preliminary principal component analysis of satellite imagery from other glaciers at other times yield nearly identical principal component band combinations. Indeed, for a particular glacier, if the user wanted a customised band combination, it would be reasonable to use PC1 and PC2 of a site-specific PCA. Thus, because spectra from the two chosen field sites and satellite images from others agree upon the principal components, it is reasonable to assume that Midtre Lovénbreen and Langjökull are representative glaciers for study.

Nevertheless, although the major classes of Midtre Lovénbreen and Langjökull will be spectrally representative of many glaciers, there are still surface features on other subsets of glaciers that are not considered but could and would impact transferability of results. The largest subset of glaciers will be those in colder climates, in particular those with dry snow facies, for example in Greenland and Antarctica. Other surfaces which fall outside the remit of this study include those influenced by debris cover, dust, black carbon, and snow algae. It is entirely possible that the LCs presented here will be appropriate for classification of these facies; earlier work (Boresjö Bronge & Bronge, 1999) classified snow and ice zonation in Antarctica (even including sea ice) using PCA as a guide. More work will be needed to confirm or deny this hypothesis, but that is beyond the scope of this study.

The analysis of the impact of spatial resolution on classification success raises the question of whether Midtre Lovénbreen is also spatially representative of other glaciers. Processes which impact the spatial distribution and scale of facies include the accumulation distribution, wind and avalanche redistribution, melt patterns, and local slopes, to name a few. In these regards, there is nothing that sets

Midtre Lovénbreen apart as a special glacier. By contrast, factors such as crevasses or incised supraglacial streams are glacier-specific and may have some impact on the impact of resolution on surface classification. However, such features are small on Midtre Lovénbreen and would likely manifest themselves as an important difference between high- and mid-resolution images, rather than influencing the conclusion that low-resolution images are inappropriate for facies classification. Thus, while there is no reason to suspect that Midtre Lovénbreen would have a different spatial character than other glaciers, it is recognized as a limitation of this study. Investigation of a wider study area would confirm or confine this extrapolation, so it is suggested as a potential future research direction.

Accepting the transferability of this study, efficient and effective multispectral glacier surface classification has many implications for glaciological research. The most immediate use is selection of optimal imagery for widespread measurements of AAR (and therefore ELA) as mass balance proxies (e.g. Rabatel et al., 2008; Rabatel et al., 2005; Shea et al., 2013). Some of the studies upon which the classification method developed in this thesis were used for validation of glacier melt and hydrology models (Braun et al., 2007; de Woul et al., 2006). The effective identification of wet facies on clean glaciers by a wide range of sensors predisposes this classification scheme to effective application to hydrological applications. Increased application to validate models studying glacier surface hydrology would therefore be appropriate. In addition to small mountain glaciers (e.g. Dahlke et al., 2012), there is increasing interest in water-saturated areas in Greenland, so this may be a promising research direction. In addition, energy balance models may find some overlap with hydrological modelling, driving the liquid contributions to the glacier system. This study provides another step in the direction of successfully applying such an assimilation or validation mechanism.

Other areas of research which will be impacted eventually include, as mentioned earlier, studies of climate variability, understanding of water resources, study of geomorphologic hazards, and investigation of high altitude and latitude biodiversity. These all tie back to better process-based understanding of glacier surface processes enabled by application of multispectral remote sensing. Here, two data sets were used to investigate the wider application of glacier surface classification across many platforms. For the new sensors whose properties are not fully quantified (OLI on Landsat 8 and Sentinel-2), it will be interesting to confirm the conclusions drawn here from the assumptions made here. Further principal component analysis of multispectral remote sensing imagery of glaciers would confirm the linear combinations of bands presented here for surface classification.

In addition, the transferability of the conclusions about the necessity of medium-to-high spatial resolution has wide implications. It is pragmatic and efficient to identify the lowest reasonable resolution of data to use for studying glacier surface properties. This is especially true if the techniques are going to be used across wide areas or long time series. Although not expected that the conclusions will change significantly, it would still be beneficial to confirm with ATM images from other glaciers, in other areas, and across larger areas as well. Beyond application of ATM images, coordinated campaigns allowing for intercomparison of coincident, multi-resolution data would lead to even more robust conclusions in the future.

7. Conclusion

Increasing availability of multispectral data requires that researchers know what data types are best suited for their own research questions. For glacier surface classification, the radiometric, spectral, and spatial properties of a suite of popular sensors (ATM, ASTER, MSS, TM, ETM+, OLI, Sentinel-2, and MODIS) are investigated using data sets in common to perform controlled analyses. Linear combinations for all sensors were created based on principal component analysis of *in situ* spectra. Among these sensors, spectral resolution and range or radiometric resolution were not important on their own. The most important property which determined the suitability of a multispectral imager for glacier surface classification was its radiometric range. In particular, it was found to be beneficial to have a low gain in the visible and a higher gain in the NIR.

Spatial resolution can, seemingly paradoxically, be either beneficial or detrimental to classification, depending on the fractal scale of the surface being classified. For glaciers, it was found that low resolution imagery (250 m pixels) is too coarse to represent the true complexity present on a glacier.

However, medium resolution imagery (60 m, 30 m, or 20 m) did accurately represent the results derived from high resolution airborne imagery. Nevertheless, 30 m imagery was preferable to 60 m imagery. From this, it was inferred that inhomogeneity on glaciers is significant on a scale between ~60 and 250 m.

Based upon these radiometric, spatial, and spectral requirements, the sensors emulated here that are most suitable for glacier surface classification are Landsat TM / TM+ (with gain settings of LLLH) and ASTER (low 1 or normal gain). Both Sentinel-2 and the OLI on Landsat 8 are also expected to be similarly qualified. Landsat MSS is also found to be radiometrically well-suited for glacier surface classification, but its lower spatial resolution makes it a secondary selection. However, MSS has historical imagery whereas other sensors have more recent (or future) data ranges, and therefore these imagers could be used in conjunction with each other. This demonstrates, once again, that although priority can be given to sensor capabilities, temporal resolution and data availability will still remain important considerations.

Consideration was given to the transferability of the results presented here. The result of universal physical processes, Midtre Lovénbreen and Langjökull are deemed to be representative of the classes present on many GICs, both spatially and spectrally. Future work should focus on downstream impacts, related to mass balance proxies, integrated with glacier modelling, and related to applied studies of glacier behaviour. The potential exists for further confirmation of the conclusions presented here using further data sets.

In sum, the work presented in this paper has contributed to the understanding of glaciological applications of multispectral remote sensing imagery, a field which will be sure to remain innovative and vital to glaciology for many years to come.

Acknowledgements

A. Pope was supported by the National Science Foundation Graduate Research Fellowship Programme under Grant No. DGE-1038596. Further research support came from UK Natural Environment Research Council's Field Spectroscopy Facility, ARCFAC (the European Centre for Arctic Environmental Research), Trinity College Cambridge, Sigma Xi, the Norwegian Marshall Fund, the Explorers Club, the

National Geographic Society Young Explorers Program, the Scott Polar Research Institute, the Cambridge University Geography Department, the Cambridge University Department of Anglo-Saxon, Norse, and Celtic Studies, and the Cambridge University Worts Fund. The thank the creators and development teams of the freely available software used in this research, including ImageJ (http://rsb.info.nih.gov/ij/), MultiSpec (https://engineering.purdue.edu/~biehl/MultiSpec/), QuantumGIS (http://www.qgis.org/), Plot (http://plot.micw.eu/), and Zotero.

The authors would also like to acknowledge the support of the staff of Norwegian Polar Institute's Sverdrup Station in Ny-Ålesund and Addi Hermannsson and Ice, Ltd. for logistical support on Langjökull and field assistance from Alexandra Giese, Richard Morris, and Jon Peatman.

References

Albert, T. H. (2002). Evaluation of Remote Sensing Techniques for Ice-Area Classification Applied to the Tropical Quelccaya Ice Cap, Peru. *Polar Geography*, *26*, 210–226.

Aniya, M., Naruso, H., Skvarca, P., & Casassa, G. (1996). The use of satellite and airborne imagery to invesntory outlet glaciers of the Southern Ratagonia Icefield, South America. *Photogrammetric Engineering and Remote Sensing*, *62*(12), 1361–1369.

Azimuth Systems. (2005, July). AZGCORR Users Manual. Retrieved from http://arsf.nerc.ac.uk/documents/azgcorr_v5.pdf

Baker, B. A., Warner, T. A., Conley, J. F., & McNeil, B. E. (2013). Does spatial resolution matter? A multi-scale comparison of object-based and pixel-based methods for detecting change associated with gas well drilling operations. *International Journal of Remote Sensing*, *34*(5), 1633–1651. doi:10.1080/01431161.2012.724540

Baraer, M., Mark, B. G., McKenzie, J. M., Condom, T., Bury, J., Huh, K.-I., … Rathay, S. (2011). Glacier recession and water resources in Peru's Cordillera Blanca. *Journal of Glaciology*, *58*(207), 134–150.

Barcaza, G., Aniya, M., Matsumoto, T., & Aoki, T. (2009). Satellite-derived equilibrium lines in Northern Patagonia Icefield, Chile, and their implications to glacier variations. *Arctic, Antarctic, and Alpine Research*, *41*(2), 174–182.

Barry, R. G. (2011). The cryosphere – past, present, and future: a review of the frozen water resources of the world. *Polar Geography*, *34*(4), 219–227. doi:10.1080/1088937X.2011.638146

Battersby, S. E., Hodgson, M. E., & Wang, J. (2012). Spatial Resolution Imagery Requirements for Identifying Structure Damage in a Hurricane Disaster: A Cognitive Approach. *Photogrammetric Engineering and Remote Sensing*, *78*(6), 625–635.

Benson, C. S. (1960). *Stratigraphic studies in the snow and firn of the Greenland ice sheet* (PhD). California Institute of Technology, Pasadena, California. Retrieved from http://etd.caltech.edu/etd/available/etd-03232006-104828/

Bindschadler, R., Choi, H., Wichlacz, A., Bingham, R., Bohlander, J., Brunt, K., … Young, N. (2011). Getting around Antarctica: new high-resolution mappings of the grounded and freely-floating boundaries of the Antarctic ice sheet created for the International Polar Year. *The Cryosphere*, *5*(3), 569–588.

Björnsson, H., & Pálsson, F. (2008). Icelandic Glaciers. *Jökull*, *58*, 365–386.

Bolch, T., Kulkarni, A., Kääb, A., Huggel, C., Paul, F., Cogley, J. G., … Stoffel, M. (2012). The State and Fate of Himalayan Glaciers. *Science*, *336*(6079), 310–314. doi:10.1126/science.1215828

Boresjö Bronge, L., & Bronge, C. (1999). Ice and snow-type classification in the Vestfold Hills, East Antarctica, using Landsat-TM - data and ground radiometer measurements. *International Journal of Remote Sensing*, *20*(2), 225. doi:10.1080/014311699213415

Braithwaite, R. J. (1984). Can the mass balance of a glacier be estimated from its equilibrium line altitude? *Journal of Glaciology*, *30*(106), 364–368.

Braithwaite, R. J., & Müller, F. (1978). On the parameterization of glacier equilibrium line altitude. *IAHS-AISH*, *126*, 263–271.

Braun, M., Schuler, T. V., Hock, R., Brown, I. A., & Jackson, M. (2007). Comparison of remote sensing derived glacier facies maps with distributed mass balance modelling at Engabreen, northern Norway. *Glacier Mass Balance Changes and Meltwater Discharge*, *IAHS 318*, 126–134.

Casacchia, R., Lauta, F., Salvatori, R., Cagnati, A., Valt, M., & �rb�k, J. B. (2001). Radiometric investigation of different snow covers in Svalbard. *Polar Research*, *20*(1), 13–22. doi:10.1111/j.1751-8369.2001.tb00035.x

Casey, K. A., Kääb, A., & Benn, D. I. (2012). Geochemical characterization of supraglacial debris via in situ and optical remote sensing methods: a case study in Khumbu Himalaya, Nepal. *The Cryosphere*, *6*(1), 85–100. doi:10.5194/tc-6-85-2012

Cogley, J. G., Hock, R., Rasmussen, L. A., Arendt, A. A., Bauder, A., Braithwaite, R. J., … Zemp, M. (2011). *Glossary of Glacier Mass Balance*. Paris: UNESCO-IHP.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. doi:10.1177/001316446002000104

Comber, A., Fisher, P., Brunsdon, C., & Khmag, A. (2012). Spatial analysis of remote sensing image classification accuracy. *Remote Sensing of Environment*, *127*, 237–246. doi:10.1016/j.rse.2012.09.005

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35–46. doi:10.1016/0034-4257(91)90048-B

Congalton, R. G., & Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL: Lewis Publishers.

Cuffey, K. M., & Patterson, W. S. B. (2010). *The Physics of Glaciers* (4th ed.). London: Academic Press.

Dahlke, H. E., Lyon, S. W., Stedinger, J. R., Rosqvist, G., & Jansson, P. (2012). Contrasting trends in floods for two sub-arctic catchments in northern Sweden – does glacier presence matter? *Hydrol. Earth Syst. Sci.*, *16*(7), 2123–2141. doi:10.5194/hess-16-2123-2012

De Angelis, H., Rau, F., & Skvarca, P. (2007). Snow zonation on Hielo Patagónico Sur, Southern Patagonia, derived from Landsat 5 TM data. *Global and Planetary Change*, *59*(1-4), 149–158. doi:10.1016/j.gloplacha.2006.11.032

De Woul, M., Hock, R., Braun, M., Thorsteinsson, T., Jóhansen, T., & Halldorsdottir, S. (2006). Firn layer impact on glacial runoff: a case study at Hofsjökull, Iceland. *Hydrological Processes*, *20*, 2171–2185.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., … Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, *120*, 25–36. doi:10.1016/j.rse.2011.11.026

Dumont, M., Brissaud, O., Picard, G., Schmitt, B., Gallet, J.-C., & Arnaud, Y. (2010). High-accuracy measurements of snow Bidirectional Reflectance Distribution Function at visible and NIR wavelengths – comparison with modelling results. *Atmos. Chem. Phys.*, *10*(5), 2507–2520. doi:10.5194/acp-10-2507-2010

Dumont, M., Gardelle, J., Sirguey, P., Guillot, A., Six, D., Rabatel, A., & Arnaud, Y. (2012). Linking glacier annual mass balance and glacier albedo retrieved from MODIS data. *The Cryosphere, 6*, 1527–1539. doi:10.5194/tcd-6-2363-2012

Dyurgerov, M. (1996). Substitution of long term mass balance data by measurements of one summer. *Gletscherkd. Glazial geol.*, *32*, 177–184.

Dyurgerov, M., Meier, M. F., & Bahr, D. B. (2009). A new index of glacier area change: a tool for glacier monitoring. *Journal of Glaciology*, *55*(192), 710–716.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, *80*(1), 185–201. doi:10.1016/S0034-4257(01)00295-4

Greuell, W., Kohler, J., Obleitner, F., Glowacki, P., Melvold, K., Bernsen, E., & Oerlemans, J. (2007). Assessment of interannual variations in the surface mass balance of 18 Svalbard glaciers from the Moderate Resolution Imaging Spectroradiometer/Terra albedo product. *Journal of Geophysical Research, 112*, D07105.

Greuell, W., & Oerlemans, J. (2005). Validation of AVHRR- and MODIS-derived albedos of snow and ice surfaces by means of helicopter measurements. *Journal of Glaciology*, *51*(172), 37–48. doi:10.3189/172756505781829575

Greuell, W., Reijmer, C. H., & Oerlemans, J. (2002). Narrowband-to-broadband albedo conversion for glacier ice and snow based on aircraft and near-surface measurements. *Remote Sensing of Environment*, *82*, 48–63.

Hall, D. K., Riggs, G. A., & Salomonson, V. V. (1995). Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, *54*(2), 127–140. doi:10.1016/0034-4257(95)00137-P

Heid, T., & Kääb, A. (2012). Repeat optical satellite images reveal widespread and long term decrease in land-terminating glacier speeds. *The Cryosphere*, *6*(2), 467–478. doi:10.5194/tc-6-467-2012

Helder, D. L., Karki, S., Bhatt, R., Micijevic, E., Aaron, D., & Jasinski, B. (2012). Radiometric Calibration of the Landsat MSS Sensor Series. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(6), 2380 –2399. doi:10.1109/TGRS.2011.2171351

Hendriks, J., & Pellikka, P. (2008). Semi-automatic glacier delineation from Landsat imagery over Hintereisferner in the Austrian Alps. *Zeitschrift für Gletscherkunde und Glazialgeologie*, *41*, 55–75.

Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology, 282*, 104–115.

Hopkinson, C., & Demuth, M. (2006). Using airborne lidar to assess the influence of glacier downwasting on water resources in the Canadian Rocky Mountains. *Canadian Journal of Remote Sensing*, *32*(2), 212–222.

Irons, J. R., Dwyer, J. L., & Barsi, J. A. (2012). The next Landsat satellite: The Landsat Data Continuity Mission. *Remote Sensing of Environment*, *122*, 11–21. doi:10.1016/j.rse.2011.08.026

Jacobsen, D., Milner, A. M., Brown, L. E., & Dangles, O. (2012). Biodiversity under threat in glacier-fed river systems. *Nature Climate Change*, *2*(5), 361–364. doi:10.1038/nclimate1435

Kargel, J. S., Abrams, M. J., Bishop, M. P., Bush, A., Hamilton, G., Jiskoot, H., … Wessels, R. (2005). Multispectral imaging contributions to global land ice measurements from space. *Remote Sensing of Environment*, *99*(1-2), 187–219. doi:10.1016/j.rse.2005.07.004

Knap, W. H., Reijmer, C. H., & Oerlemans, J. (1999). Narrowband to broadband conversion of Landsat TM glacier albedos. *International Journal of Remote Sensing*, *20*(10), 2091–2110.

Kokhanovsky, A. A., & Zege, E. P. (2004). Scattering Optics of Snow. *Applied Optics*, *43*(7), 1589–1602. doi:10.1364/AO.43.001589

König, M., Winther, J.-G., Kohler, J., & König, F. (2004). Two methods for firn-area and mass-balance monitoring of Svalbard glacier with SAR satellite images. *Journal of Glaciology*, *50*(168), 116–128.

Kuchiki, K., Aoki, T., Niwano, M., Motoyoshi, H., & Iwabuchi, H. (2011). Effect of sastrugi on snow bidirectional reflectance and its application to MODIS data. *Journal of Geophysical Research*, *116*, 15 PP. doi:201110.1029/2011JD016070

Li, W., Stamnes, K., Chen, B., & Xiong, X. (2001). Snow grain size retrieved from near- infrared radiances at multiple wavelengths. *Geophysical Research Letters*, *28*(9), PAGES 1699–1702.

Luckman, A., Jansen, D., Kulessa, B., King, E. C., Sammonds, P., & Benn, D. I. (2012). Basal crevasses in Larsen C Ice Shelf and implications for their global abundance. *The Cryosphere*, *6*(1), 113–123. doi:10.5194/tc-6-113-2012

Machguth, H., Paul, F., Kotlarski, S., & Hoelzle, M. (2009). Calculating distributed glacier mass balance for the Swiss Alps from regional climate model output: A methodical description and interpretation of the results. *Journal of Geophysical Research*, *114*(D19), D19106. doi:10.1029/2009JD011775

Markham, B. L., & Helder, D. L. (2012). Forty-year calibrated record of earth-reflected radiance from Landsat: A review Keywords: Landsat, Radiometric Calibration, History. MSS, TM, ETM+. *Remote Sensing of Environment*, *122*, 30–40. doi:10.1016/j.rse.2011.06.026

Michishita, R., Jiang, Z., Gong, P., & Xu, B. (2012). Bi-scale analysis of multitemporal land cover fractions for wetland vegetation mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, *72*, 1–15.

Monserud, R., & Leemans, R. (1992). Comparing Global Vegetation Maps with the Kappa-Statistic. *Ecological Modelling*, *62*(4), 275–293. doi:10.1016/0304-3800(92)90003-W

NASA. (2002, August). ASTER User Handbook, Version 2. Retrieved from http://asterweb.jpl.nasa.gov/content/03_data/04_Documents/aster_user_guide_v2.pdf

NASA. (2011). Landsat 7 Science Data Users Handbook. National Aeronautics and Space Administration. Retrieved from http://landsathandbook.gsfc.nasa.gov/

Nolin, A. W., & Dozier, J. (1993). Estimating snow grain size using AVIRIS data. *Remote Sensing of Environment*, *44*(2-3), 231–238. doi:10.1016/0034-4257(93)90018-S

Nolin, A. W., & Payne, M. C. (2007). Classification of glacier zones in western Greenland using albedo and surface roughness from the Multi-angle Imaging SpectroRadiometer (MISR). *Remote Sensing of Environment*, *107*(1-2), 264–275. doi:10.1016/j.rse.2006.11.004

Oerlemans, J. (1994). Quantifying Global Warming from the Retreat of Glaciers. *Science*, *264*(5156), 243–245. doi:10.1126/science.264.5156.243

Orheim, O., & Lucchitta, B. (1987). Snow and ice studies by thematic mapper and multispectral scanner landsat images. *Annals of Glaciology*, *9*, 109–118.

Painter, T. H. (2011). Comment on Singh an others, "Hyperspectral analysis of snow reflectance to understand the effects of contamination and grain size". *Journal of Glaciology*, *57*(201), 183–185.

Paul, F. (2000). Evaluation of different methods for glacier mapping using Landsat TM. In *EARSeL-SIG-Workshop Land Ice and Snow* (Vol. 1, pp. 238–245). Desden/FRG.

Paul, F., Barrand, N. E., Berthier, E., Bolch, T., Casey, K. A., Frey, H., … Winsvold, S. (2013). On the accuracy of glacier outlines derived from remote sensing data. *Annals of Glaciology*, *54*(63), 171–182.

Paul, F., Huggel, C., Kääb, A., Kellenberger, T., & Maisch, M. (2002). Comparison of TM-Derived Glacier Areas with Higher Resolution Data Sets. *EARSeL eProceedings*, *2*, 15–21.

Paul, F., & Kääb, A. (2005). Perspectives on the production of a glacier inventory from multispectral satellite data in Arctic Canada: Cumberland Peninsula, Baffin Island. *Annals of Glaciology*, *42*, 59–66. doi:10.3189/172756405781813087

Pellikka, P., & Rees, W. G. (2010). *Remote Sensing of Glaciers*. London: CRC Press.

Phinn, S. R., Roelfsema, C. M., & Mumby, P. J. (2012). Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs. *International Journal of Remote Sensing*, *33*(12), 3768–3797. doi:10.1080/01431161.2011.633122

Pope, A., & Rees, W. G. (In Press). Using in situ Spectra to Explore Landsat Classification of Glacier Surfaces. *Journal of Applied Earth Observation and Geoinformation*.

Pope, A., Willis, I. C., Rees, W. G., Arnold, N. S., & Pálsson, F. (2013). Combining airborne lidar and Landsat ETM+ data with photoclinometry to produce a digital elevation model for Langjökull, Iceland. *International Journal of Remote Sensing*, *34*(4), 1005–1025. doi:10.1080/01431161.2012.705446

Rabatel, A., Dedieu, J.-P., Thibert, E., Letréguilly, A., & Vincent, C. (2008). 25 years (1981-2005) of equilibrium-line altitude and mass-balance reconstruction on Glacier Blanc, French Alps, using remote-sensing methods and meteorological data. *Journal of Glaciology*, *54*, 307–314. doi:10.3189/002214308784886063

Rabatel, A., Dedieu, J.-P., & Vincent, C. (2005). Using remote-sensing data to determine equilibrium-line altitude and mass-balance time series: validation on three French glaciers, 19942002. *Journal of Glaciology*, *51*, 539–546. doi:10.3189/172756505781829106

Rees, W. G. (2006). *Remote Sensing of Snow and Ice*. Boca Raton: Taylor & Francis.

Rees, W. G. (2008). Comparing the spatial content of thematic maps. *International Journal of Remote Sensing*, *29*(13), 3833–3844.

Rees, W. G., & Arnold, N. S. (2007). Mass Balance and Dynamics of a Valley Glacier Measured by High-Resolution LiDAR. *Polar Record*, *43*(04), 311–319. doi:10.1017/S0032247407006419

Reijmer, C. H., Knap, W. H., & Oerlemans, J. (1999). The Surface Albedo of Vatnajokull Ice Cap, Iceland: A Comparison Between Satellite-Derived and Ground-Based Measurements. *Boundary-Layer Meteorology*, *92*, 125–144.

Scambos, T. A., Frezzotti, M., Haran, T., Bohlander, J., Lenaerts, J. T. M., Van den Broeke, M., … Winther, J.-G. (2012). Extent of low-accumulation “wind glaze” areas on the East Antarctic plateau: implications for continental ice mass balance. *Journal of Glaciology*, *58*(210), 633–647. doi:10.3189/2012JoG11J232

Schaepman-Strub, G., Schaepman, M. E., Painter, T. H., Dangel, S., & Martonchik, J. V. (2006). Reflectance quantities in optical remote sensing - definitions and case studies. *Remote Sensing of Environment*, *103*(1), 27–42. doi:10.1016/j.rse.2006.03.002

Shea, J. M., Menounos, B., Moore, R. D., & Tennant, C. (2013). An approach to derive regional snow lines and glacier mass change from MODIS imagery, western North America. *The Cryosphere*, *7*(2), 667–680. doi:10.5194/tc-7-667-2013

Shukla, A., Gupta, R. P., & Arora, M. (2009). Estimation of debris cover and its temporal variation using optical satellite sensor data: a case study in Chenab basin, Himalaya. *Journal of Glaciology*, *55*(191), 444–452.

Sidjak, R. W., & Wheate, R. D. (1999). Glacier mapping of the Illecillewaet icefield, British Columbia, Canada, using Landsat TM and digital elevation data. *International Journal of Remote Sensing*, *20*(2), 273–284.

Sobrino, J. A., Oltra-Carrió, R., Sòria, G., Bianchi, R., & Paganini, M. (2012). Impact of spatial resolution and satellite overpass time on evaluation of the surface urban heat island effects. *Remote Sensing of Environment*, *117*, 50–56. doi:10.1016/j.rse.2011.04.042

Takeuchi, N. (2009). Temporal and spatial variations in spectral reflectance and characteristics of surface dust on Gulkana Glacier, Alaska Range. *Journal of Glaciology*, *55*(192), 701–709.

Van Angelen, J. H., Lenaerts, J. T. M., Lhermitte, S., Fettweis, X., Kuipers Munneke, P., Van den Broeke, M. R., … Smeets, C. J. P. P. (2012). Sensitivity of Greenland Ice Sheet surface mass balance to surface albedo parameterization: a study with a regional climate model. *The Cryosphere*, *6*(5), 1175–1186. doi:10.5194/tc-6-1175-2012

Williams, R. S., Hall, D. K., & Benson, C. S. (1991). Analysis of glacier facies using satellite techniques. *Journal of Glaciology*, *37*(125), 120–128.

Wiscombe, W. J., & Warren, S. G. (1980). A Model for the Spectral Albedo of Snow. I: Pure Snow. *Journal of the Atmospheric Sciences*, *37*(12), 2712–2733.

Wolken, G. J., Sharp, M. J., & Wang, L. (2009). Snow and ice facies variability and ice layer formation on Canadian Arctic ice caps, 1999–2005. *Journal of Geophysical Research, 114*, F03011. doi:10.1029/2008JF001173

Zeng, Q., Cao, M., Feng, X., Liang, F., Chen, X., & Sheng, W. (1983). A study of spectral reflection characteristics for snow, ice and water in the north of China. *Hydrological Applications of Remote Sensing and Remote Data Transmissions*, *145*, 451–462.

Figures:

Figure 1: Comparison of spectral bands of the multispectral sensors considered in this study. Background spectrum is the same as 'Fine Snow' in Figure 2.3. See Table S2 for further details.

Figure 2: False-colour ATM image of Midtre Lovénbreen collected on 9 August, 2003. ATM bands 4, 3, and 2 are used for red, green, and blue respectively. Full colour versions for this and other figures are available online.

Figure 3: Biplots of the first and second linear combinations of *in situ* spectral data from Midtre Lovénbreen modified to mimic the spectral and radiometric properties of a range of multispectral imagers. See Table 1 for the details of all individual plots, (a) through (p). The ellipses in Figure 3a also show the three groups into which the data were classified.

Figure 4: Biplots of the first and second linear combinations of *in situ* spectral data from Langjökull modified to mimic the spectral and radiometric properties of a range of multispectral imagers. See Table 2 for the details of all individual plots, (a) through (p). The ellipse in Figure 4a highlights the clean ice spectra which were classified as distinct from the rest.

Figure 5: Midtre Lovénbreen surface classification using ATM imagery at 2 m resolution (a) and the same image degraded to 20 m (b), 30 m (c), 60 m (d), and 250 m pixels (e).

Tables:

Table 1: For Midtre Lovénbreen spectra: sensors, radiometric resolution, gain settings, *K* (Cohen's Kappa) of clustering success, description of classification success (Monserud & Leemans, 1992), and corresponding panel in Figure 3.

Sensor	Radiometric Resolution	Gain Settings	K	Monserud Agreement	Figure 6.1
ASTER	8 bit	hi	0.9816	Excellent	b
ASD FieldSpec	8 bit	-	0.9742	Excellent	c
ASD FieldSpec	16 bit	-	0.9742	Excellent	d
ASD FieldSpec	12 bit	-	0.9742	Excellent	e
Landsat MSS	8 bit	-	0.9705	Excellent	f
ASTER	8 bit	low 1	0.9669	Excellent	g
ASD FieldSpec	-	-	0.9669	Excellent	a
ATM	16 bit	-	0.9631	Excellent	h
ASTER	8 bit	normal	0.9595	Excellent	i
Sentinel-2	12 bit	-	0.9484	Excellent	j
Landsat TM / ETM+	8 bit	LLLH	0.9448	Excellent	k
Landsat TM / ETM+	8 bit	HHHH	0.9267	Excellent	l
MODIS 1-7	12 bit	-	0.9158	Excellent	m
MODIS 8+	12 bit	hi	0.8938	Excellent	n
MODIS 8+	12 bit	lo	0.8721	Excellent	o
Landsat TM / ETM+	8 bit	LLLL	0.7977	Very Good	p

Table 2: For Langjökull spectra: sensors, radiometric resolution, gain settings, K (Cohen's Kappa) of clustering success, description of classification success (Monserud & Leemans, 1992), and corresponding panel in Figure 4.

Sensor	**Radiometric Resolution**	**Gain Settings**	***K***	**Monserud Agreement**	**Figure 6.2**
ASD FieldSpec	-	-	1	Perfect	a
ASD FieldSpec	16 bit	-	1	Perfect	b
ASD FieldSpec	12 bit	-	1	Perfect	c
ASD FieldSpec	8 bit	-	1	Perfect	d
Sentinel-2	12 bit	-	1	Perfect	e
Landsat TM / ETM+	8 bit	LLLH	1	Perfect	f
ASTER	8 bit	low 1	1	Perfect	g
Landsat MSS	8 bit	-	1	Perfect	h
ASTER	8 bit	normal	0.9081	Excellent	i
MODIS 1-7	12 bit	-	0.6579	Good	j
Landsat TM / ETM+	8 bit	HHHH	0.6559	Good	k
ATM	16 bit	-	0.5979	Good	l
Landsat TM / ETM+	8 bit	LLLL	0.5841	Good	m
ASTER	8 bit	hi	0.5101	Fair	n
MODIS 8+	12 bit	hi	0.4879	Fair	o
MODIS 8+	12 bit	lo	0.3522	Poor	p

Table 3: Comparison of the results of the classifications presented in Figure 5 by considering relative area of each class as well as their aggregation into glaciologically meaningful groups of accumulation and ablation area.

Class	2 m Pixels	Percent	Interpretation			
1	124,425	9.7	Shadow / Debris			
2	196,744	15.3	Shadow / Debris			
3	57,694	4.5	Ablation			
4	67,947	5.3	Ablation		Total Area:	5.15 km^2
5	107,323	8.3	Ablation			
6	158,711	12.3	Ablation		Shadow/Debris:	25.0 %
7	205,371	16.0	Ablation		Ablation:	60.4 %
8	179,783	13.0	Ablation		Accumulation:	14.7 %
9	116,938	9.1	Accumulation			
10	72,040	5.6	Accumulation			
Class	**20 m Pixels**	**Percent**	**Interpretation**			
1	1,383	10.5	Shadow / Debris			
2	1,932	14.7	Shadow / Debris			
3	610	4.6	Ablation			
4	653	5.0	Ablation		Total Area:	5.26 km^2
5	1,023	7.8	Ablation			
6	1,479	11.3	Ablation		Shadow/Debris:	25.0 %
7	2,307	17.6	Ablation		Ablation:	60.5 %
8	1,883	14.3	Ablation		Accumulation:	14.2 %
9	1,107	8.4	Accumulation			
10	765	5.8	Accumulation			
Class	**30 m Pixels**	**Percent**	**Interpretation**			
1	630	10.6	Shadow / Debris			
2	889	15.0	Shadow / Debris			
3	287	4.8	Ablation			
4	288	4.9	Ablation		Total Area:	5.33 km^2
5	454	7.7	Ablation			
6	624	10.5	Ablation		Shadow/Debris:	25.7 %
7	1,005	17.0	Ablation		Ablation:	60.0 %
8	895	15.1	Ablation		Accumulation:	14.3 %
9	491	8.3	Accumulation			
10	358	6.0	Accumulation			

Class	60 m Pixels	Percent	Interpretation			
1	167	10.8	Shadow / Debris			
2	228	14.8	Shadow / Debris			
3	93	6.0	Ablation			
4	87	5.6	Ablation		Total Area:	5.55 km^2
5	92	6.0	Ablation			
6	153	9.9	Ablation		Shadow/Debris:	25.6 %
7	240	15.6	Ablation		Ablation:	60.2 %
8	264	17.1	Ablation		Accumulation:	14.2 %
9	119	7.7	Accumulation			
10	100	6.5	Accumulation			
Class	**250 m Pixels**	**Percent**	**Interpretation**			
1	15	13.4	Shadow / Debris			
2	16	14.3	Shadow / Debris			
3	7	6.3	Ablation			
4	11	9.8	Ablation		Total Area:	7.0 km^2
5	7	6.3	Ablation			
6	6	5.4	Ablation		Shadow/Debris:	27.7 %
7	13	11.6	Ablation		Ablation:	54.5 %
8	17	15.2	Ablation		Accumulation:	17.9 %
9	11	9.8	Accumulation			
10	9	8.0	Accumulation			

Table 4: Comparison of the results of the classifications presented in Figure 5 by down-sampling results with majority filtering to lower resolutions. A is overall agreement and K is Cohen's Kappa, both presented in Section 4.5; perfect agreement would lead to a value of 1 for both A and K.

Starting Resolution	Final Resolution	*A*	*A*, grouped	*A*, ranked	*K*	*K*, grouped	*K*, ranked
2 m	30 m	0.6315	0.9477	1	0.5825	0.9035	1
2 m	20 m	0.7217	0.9459	2	0.6837	0.9017	2
20 m	30 m	0.5788	0.9324	3	0.5226	0.8746	3
30 m	60 m	0.6414	0.9084	6	0.5927	0.8360	4
2 m	60 m	0.5284	0.9135	4	0.4662	0.8359	5
20 m	60 m	0.6126	0.9090	5	0.5605	0.8326	6
60 m	250 m	0.2857	0.7727	7	0.1881	0.5956	7
30 m	250 m	0.1923	0.7612	8	0.0878	0.5681	8
20 m	250 m	0.1975	0.6857	10	0.1023	0.4554	9
2 m	250 m	0.2405	0.6912	9	0.1425	0.4322	10

Table 5: Comparison of the results of the classifications presented in Figure 5 by downscaling results to higher resolutions. A is overall accuracy agreement and K is Cohen's Kappa, both described in Section 4.5; perfect agreement would lead to a value of 1 for both A and K.

Starting Resolution	**Final Resolution**	***A***	***A*, grouped**	***A*, ranked**	***K***	***K*, grouped**	***K*, ranked**
30 m	20 m	0.7381	0.9603	1	0.7030	0.9269	1
20 m	2 m	0.6393	0.9585	2	0.5918	0.9233	2
30 m	3 m	0.5976	0.9450	3	0.5443	0.8983	3
60 m	30 m	0.6383	0.9338	4	0.5902	0.8778	4
60 m	20 m	0.5958	0.9252	5	0.5422	0.8609	5
60 m	2 m	0.4977	0.9065	6	0.4317	0.8253	6
250 m	60 m	0.2612	0.6909	7	0.1690	0.4469	7
250 m	30 m	0.2342	0.6833	8	0.1400	0.4223	8
250 m	20 m	0.2165	0.6711	9	0.1207	0.3985	9
250 m	2 m	0.2145	0.6647	10	0.1190	0.3883	10