A decision-tree classification for low-lying complex land cover types within the zone of discontinuous permafrost

L. Chasmer a,c,⁎, C. Hopkinson c, T. Veness b, W. Quinton b, J. Baltzer d

a Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada
b Cold Regions Research Centre, Wilfrid Laurier University, Waterloo, ON N2L 3C1, Canada
c Department of Geography, University of Lethbridge, Lethbridge, AB T1K 3M4, Canada
d Department of Biology, Wilfrid Laurier University, Waterloo, ON N2L 3C1, Canada

A R T I C L E   I N F O
Article history:
Received 8 August 2013
Received in revised form 20 December 2013
Accepted 23 December 2013
Available online xxxx

Keywords:
Discontinuous permafrost
Peatland
Classification
Airborne LiDAR
WorldView 2
Discharge
Permafrost thaw
Hydrological model

A B S T R A C T
This study presents a decision-tree (DT) approach to classifying heterogeneous land cover types within a northern watershed located in the zone of discontinuous permafrost using airborne LiDAR and high resolution spectral datasets. Results are compared with a more typically applied supervised classification. Increasing errors in discharge resulting from an inaccurate classification are quantified using a distributed hydrological model. The hierarchical classification was accurate between 88% and 97% of the validation sub-area, whereas the parallel-epiped classification was accurate between 38% and 74% of the same area (despite overall accuracy of ~91%, kappa = 0.91). Topographical derivatives were best able to explain variations in land cover types (82% to 96%), whilst spectral and vegetation structural derivatives were less accurate. When compared with field measurements, the hierarchical classification of plateau edges (adjacent to a fen) was within 2 m of measured, 60% of the time, whilst this occurred only 40% of the time when using a spectral classification. When examining the impacts of land cover classification accuracy on modelled discharge, we find that the length of the Hydrological Response Unit defined by the classification (and subject to varying levels of errors) was linearly related to discharge (m³/s) such that an increase in permafrost plateau area would increase discharge by 26% of the total. The methodology presented in this paper clarifies previous classification and modelling studies using Landsat and IKONOS data for the same basin. This study greatly improves upon past classifications in the same area, furthers our understanding of the distribution of connected bogs and fens (as conveyors of water to the basin outlet) within the watershed, and current spatial extents of rapidly thawing permafrost plateaus, which are critical for better understanding the impacts of climate change on these northern environments.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction
The zone of discontinuous permafrost has undergone significant climate warming and permafrost loss over the past number of decades (e.g. Anisimov & Reneva, 2006; Beilman, Vitt, & Halsey, 2001; Chasmer, Hopkinson, & Quinton, 2010; Quinton, Hayashi, & Chasmer, 2010; Shur & Jorgenson, 2007). This is especially evident in colder permafrost areas, which are subject to rapid permafrost warming (Romanovsky, Smith, & Christiansen, 2010) and the existence of thermal inertia within warm, thin perennially frozen ground (Lewkowicz, Etzelmüller, & Smith, 2011). Permafrost underlies approximately 25% of the total land area within the Northern Hemisphere and therefore small shifts in ground heating and vegetation succession and the associated changes in permafrost distribution and extent can have globally relevant implications (e.g. Jorgenson, Racine, Walters, & Osterkamp, 2001). In the discontinuous permafrost zone, rates of permafrost thaw are expected to accelerate (Anisimov & Reneva, 2006) as plateaus become increasingly fragmented (Baltzer, Veness, Chasmer, & Quinton, in press; Chasmer et al., 2010). This can have significant impacts on both human and environmental systems, including greenhouse gas fluxes (Chasmer, Kenward, Quinton, & Petrone, 2012; Myers-Smith, McGuire, Harden, & Chapin, 2007), forest fires (Camill & Clark, 2000); changes to surface hydrology and flooding (Guan, Westbrook, & Spence, 2010; Wright, Hayashi, & Quinton, 2009); and northern infrastructure and economy (Prowse et al., 2009).

Accurate classification of the spatial distribution of land cover types, especially in areas that are rapidly changing (e.g. Chasmer et al., 2012), is fundamentally important for quantifying how these changes are affecting ecosystems (Foody, 2002). Land cover change, often as a result of climatic or anthropogenic disturbance, is viewed as the single most important variable affecting ecosystem processes (e.g. Foody, 2002; Vitousek, 1994), whilst our ability to predict future global environmental
scenarios as a result of climate change depends significantly on the accuracy of land cover classification (e.g. Feddema et al., 2005). Remote sensing data are most frequently used for the classification of land cover types (e.g. Heginbottom, 2002). Data are spatially continuous and provide a recognizable photographic appearance of the Earth’s surface, thus facilitating the comparison of features of interest through space and/or time. Furthermore, datasets often have a lengthy history of acquisition, which can be used for the detection of land cover change or conditions through time (e.g. Heginbottom, 2002).

Remote sensing-based classification of permafrost extent and ice content over broad areas has had early and ongoing interest, especially where multiple layers of land surface characteristics (e.g. vegetation, topography, etc.) correlating to the existence of permafrost are used. Early studies attempted to classify ranges of active layer thickness using thermal imagery and visible layers of vegetation cover (Morrissette, Strong, & Card, 1986; Peddle & Franklin, 1993) and topographic derivatives (Peddle & Franklin, 1993), but were unable to quantify useful ranges relative to in situ spring-time measurements. Vitt, Halsey, and Zoltai (1994) visually assessed aerial photographs acquired in Alberta, Saskatchewan, and Manitoba between 1949 and 1953 and counted the numbers of bogs within each photograph, manually assigning rare to abundant classes to each. They found that permafrost areas containing bogs had rates of degradation that were greater than rates of aggradation. By the late 1990s, more sophisticated methods of automated classification (e.g. neural networks) were applied to correlating indicators of permafrost (e.g. Leverington & Duguay, 1997), and whilst less time intensive, these were not easily transferred between sites. The integration of remote sensing data within statistical models of mountain permafrost distribution (Gruber & Hoelzle, 2001) did not improve model accuracy of permafrost prediction, possibly due to complex non-linear feedbacks between energy inputs to the surface and permafrost losses in some areas but not others (e.g. Anisimov & Renova, 2006). Early permafrost classification accuracies ranged from approximately 40% to 70%.

Spectral classifications of land cover types identified using remote sensing data within the zone of discontinuous permafrost are often problematic due to fragmented land cover boundaries, low spectral contrast between land cover types, and rapidly changing spectral characteristics at bog/fen and plateau edges as a result of soil moisture changes. Further, the extension of often living but unhealthy “remnant” trees beyond plateau boundaries makes it exceedingly difficult to accurately represent true plateau edges using spectral data alone (Chasmer et al., 2010). This is important as historical rates of permafrost and land cover changes become an indicator of the effects of climate change on northern environments. In some cases of permafrost change detection, the accuracy of the classification was not discussed. Classification and geometric errors may lead to substantial inaccuracies in permafrost extent, which will propagate uncertainties associated with the quantification of permafrost loss/land cover change as well as the use of land cover maps within land surface and hydrological models (e.g. Miller, Guertin, & Goodrich, 2007). Studies that combined digital elevation models with spectral image classification and multi-temporal data were better able to characterise permafrost by minimising misclassification errors. Nguyen, Burn, King, and Smith (2009) were able to map permafrost extent to approximately 90% accuracy using high resolution SPOT imagery of vegetation communities through the use of spectral vegetation indices, texture analysis, and principal components analysis (PCA). However greater than 90% of the land surface was underlain by permafrost, and the application of the methodology to more heterogeneous (discontinuous permafrost) regions was not assessed. Other methods, including object-based image analysis (e.g. Hay, Blaschke, Marceau, & Bouchard, 2003; Johansen, Coops, Gergel, & Stange, 2007) that use the pixel spectrum, spatial location, spectral homogeneity, and clustered shapes to identify objects can be highly accurate, but also require user intervention which may or may not be applicable to broad areas of discontinuous permafrost. New applications of Random Forest and machine learning classification methods have been applied with high accuracies to spectral remote sensing data (e.g. Rodriguez-Galiano, Ghimire, Rogan, Chica-Olmo, & Rigol-Sanchez, 2012) and LiDAR data (e.g. Im, Jensen, & Hodgson, 2008) but not in the zone of discontinuous permafrost.

Land cover classifications integrating airborne LiDAR data with spectral remote sensing data for characterising vegetated environments are fewer in number than traditional spectral classifications, but are gaining popularity. Several studies have found marked improvements via the integration of textural, topographic and vegetation structure characteristics as well as spectral data in other regions (e.g. Goodale, Hopkinson, Colville, & Amirault-Langlais, 2007). Use of digital elevation models (DEMs) of underlying topography, LiDAR data products, and LiDAR/spatial data fusion classification methods have also become popular in mountainous permafrost areas (e.g. Kremer, Lewkowicz, Bonnavourte, & Sawada, 2011) where geomorphic changes due to permafrost thaw pose a considerable hazard. In the zone of discontinuous permafrost, use of LiDAR has been less popular, likely due to logistical expenses; however, there are a few notable studies. LiDAR data were used to map the existence of permafrost based on land cover characteristics (Panda, Prakash, Solie, Romanovsky, & Jorgenson, 2010) and wet areas from laser pulse intensity (Stevens & Wolfe, 2012). Research by Hubbard et al. (2012) integrated LiDAR data with geophysical data to characterise above- and below-ground linkages between permafrost, land surface properties, and sub-surface hydrology/energy balance. We have not yet found a study that integrates airborne LiDAR and spectral data fusion methods for characterising land cover classes within the zone of discontinuous permafrost. This is currently a highly relevant and topical area of research required for better understanding the sensitivity of these northern ecosystems to development, resources extraction, and (natural/anthropogenic) disturbance.

In this study we present a decision-tree land cover classification methodology for permafrost plateaus, bogs, fens, uplands and water (ponds, lakes). The classification combines multiple-resolution spectral, textural, and three-dimensional sub-tree classification layers within the global decision hierarchy. Sensitivity analysis is used to determine the greatest contributors to identification of land cover types, with comparisons to a supervised land cover classification of spectral WorldView 2 data. Classification accuracies are evaluated against field measurements, and implications of the classification accuracy are illustrated using a hydrological runoff model.

2. Study area

The Scotty Creek watershed (61.44°N, 121.25°W) is located ~50 km south of Fort Simpson within the zone of sporadic discontinuous permafrost (Hegginbottom, Dubreil, & Harker, 1995), Northwest Territories, Canada (Fig. 1). The ~150 km² watershed is comprised of a highly heterogeneous mosaic of small permafrost mounds ~100 m² (palsa; Beilman et al., 2001) and larger plateaus (up to 20,000 m²), ombrotrophic flat bogs, channel fens, upland moraine deposits with a dense cover of deciduous and/or spruce trees, and small lakes and ponds. Permafrost thickness varies with ground cover and ranges from a few metres to over 20 m (Smith, Burgess, & Riseborough, 2008), but in general, is very thin and warm (Smith et al., 2008). Plateau coverage has been estimated at ~22% for the year 2000 with a predicted reduction to ~17% by 2055 (Duchesne, Wright, & Ednie, 2008), whilst Beilman and Robinson (2003) have shown losses of 22% on average over the past 50 years in this area. Permafrost is typically found in organic terrain and was likely formed during the Little Ice Age and therefore, it is not in equilibrium with the current climate (Shur & Jorgenson, 2007). Further, permafrost plateaus and palsas, which rise slightly above peatlands, are surrounded by unfrozen and often very wet ground. This is some thermal influence on the degradation of permafrost at plateau edges (e.g. Quinton et al., 2010). Well drained upland moraine deposits covered with shallow
organic soils and predominantly deciduous or mixed forest are believed to be largely absent of permafrost.

From a remote sensing perspective, permafrost plateaus are characterised by spectrally “dark” (visible wavelengths) mature black spruce (Picea mariana), which vary in height between 2 m and 12 m, and with effective LAI ranging between 1.1 and 1.4 m² m⁻² (on average). Ground cover vegetation, common Labrador tea (Rhododendron groenlandicum), bog cranberry (Vaccinium vitis-idaea), rusty peat moss (Sphagnum fuscum), yellow reindeer lichen (Cladonia mitsi) and grey reindeer lichen (Cladonia rangifer) sit on a newly decomposed fibric peat layer (0.2–0.5 m) under which an organic peat layer extends to a depth of up to 8 m (Quinton, Hayashi, & Pietroniro, 2003). Channel fans are easily identified as large linear features which convey water received from permafrost to basin outlets (usually ponds) (Quinton et al., 2003). Fen surfaces are covered by a buoyant peat mat of various pleurocarp mosses including golden fuzzy fen moss (Tomentypnum nitens), tufted moss (Aulacomnium palustre) and stick hook moss (Hamatocaulis vernicosus). The vascular vegetation is dominated by sedges, herbs such as buck-bean (Menyanthes trifoliolina) and shrubs including birches (Betula spp), dwarf bog-rosemary (A. polifolia) and sweet gale (Myrica gale). The peat mat sits just above the water table and responds to fluctuations in water table height, but is not able to support tall tree development due to ground surface instability (except for sparse larch (Larix laricina) in some parts). Fen soils are comprised of a dense organic layer with some mineral soils, which extend to a depth of 3 m below the water surface (Hayashi, Quinton, Pietroniro, & Gibson, 2004). Bog surfaces are highly reflective in visible and infrared wavelengths due to a cover of various species of peat moss, with some ericaceous shrubs: leather leaf (Chamaedaphne calyculata), dwarf bog-rosemary (A. polifolia) and small bog cranberry (Vaccinium oxyccos) as well as some herbs including Maianthemum trifolium. Bogs are often spectrally confused with fens and plateaus in areas where the water table is at the ground surface, yet often small (<100 m²) bog patterns with rounded edges, surrounded by upraised permafrost plateaus make them morphologically unique to fens. Moraine uplands are comprised of dense trembling aspen (Populus tremuloides), white spruce (Picea glauca), and Alaskan birch (Betula nealaska). Tree species are much taller with greater foliage cover than that found in other parts of the watershed due to rocky, mineral rich soils.

3. Materials and methods

3.1. Field data collection

Ten transects traversing fen, plateau and bog land cover types (edges of uplands and lakes were not measured due to inaccessibility) were established throughout the growing season 2011, one year following the LiDAR data collection. Environmental measurements including snow depth, top of soil profile soil moisture (0 to 5 cm), depth to frost table (at first refusal using a graduated steel rod), vegetation species type, and canopy gap fraction (estimated using digital hemispherical photography) were geographically located using a Leica (Leica Geosystems Inc., Canada) SR530 RTK (real-time kinematic) differential GPS system (Baltzer et al., in press). Land cover type was also visually identified according to similarity of characteristics representing what makes up each land cover type. Because land cover type only was used for validation, and due to the spacing of the GPS measurement locations of depth to frost table (−5 m), it was assumed that land cover beyond plateau edges would not change drastically within one year.

The boundary between the fen and permafrost plateau was defined as the line created where the water surface met the sloped edge of the permafrost plateau. This plateau ‘edge’ was also surveyed at the same time and same day as the airborne LiDAR data collection on August 2nd, 2010. The waterline was visually identified at intervals ranging between 2 and 13 m around the edge of a single large plateau, and surveyed using a Leica SR530 (Leica Geosystems Inc., Canada) differential GPS system in post-process kinematic (PPK) mode (system accuracy = +/−0.02 m). The GPS base station was set up in an open area in the middle of the plateau within 200 m of all measurements. The same base station data were also used to differentially correct the airborne LiDAR trajectory so there is high confidence in the spatial co-registration of the field and airborne data products.

3.2. Remote sensing data collection and processing

Vegetation structural characteristics and topographic derivatives used in the hierarchical land cover classification were derived from airborne scanning LiDAR data, planned and collected by the authors on August 2, 2010 using an Optech Inc. (Toronto, Ontario) ALTM3100c (Hopkinson et al., 2013). The sensor was operated at a flying height of 1500 m.a.g.l. with a pulse repetition frequency of 50 kHz, a scan angle of ±20° and 50% overlap of scan lines. The approximate number of
returns per m² is 2. Data processing following initial point cloud integration with GPS and internal measurement unit (IMU) data, and calibration included the removal of isolated and erroneous returns, ground classification, flightline alignment and ground classification smoothing (to minimise random errors to a tolerance of 0.05 m) (TerraScan, TerraSolid Inc., Finland). LiDAR-based data derivatives used as initial inputs into the classification of land cover types are summarised in Fig. 2.

WorldView 2 (Digital Globe Corp.) data were acquired at 20° maximum off nadir on October 1st, 2010. Image bands included a 0.65 m resolution panchromatic image (400–900 nm), and eight narrow spectral bands: coastal (400–450 nm), blue (450–510 nm), green (510–580 nm), yellow (585–625 nm), red (630–690 nm), red edge (705–745 nm), near infrared 1 (NIR1, 770–895 nm), and near infrared 2 (NIR2, 860–1040 nm) at 2 m pixel resolution. Mosaicking and colour balancing was performed in Geomatica OrthoEngine (PCI Inc. Canada) and orthorectified using the LiDAR DEM, 2 m resolution gridded ground surface laser return intensity and 23 tie points located within both datasets at the intersections of trails and seismic lines. Ortho accuracy was better than 1 m using a second order polynomial transformation in OrthoEngine. Pixel digital numbers (DN) were then converted to top of atmosphere spectral radiance (W m⁻² sr⁻¹ μm⁻¹) based on absolute radiometric calibration factors (W m⁻² sr⁻¹ count⁻¹) provided per band to get a band-integrated radiance (W m⁻² sr⁻¹). This is then divided by effective bandwidth to get spectral radiance combined with DEM topography and used within the ATCOR 2 atmospheric correction module of Geomatica (Richter, 2009) to remove minimal haze and variable optical depths, for colour balancing and to calculate surface reflectance.  

3.3. Classification methodology

In this study we use a decision-tree (DT) fusion classifier to quantify the highest probability of prediction given a set of input data and 50 training sites per land cover class, distributed throughout the watershed. Hierarchical models are based on a set of decisions and rules that are applied at each node of the classifier depending on the probability of the input to accurately predict the land cover of the training set (Zhu, Liu, & Jia, 2006). Primary inputs (e.g. DEM) and secondary derivatives (e.g. slope, aspect) datasets are beneficial for classification because they not only characterise spectral differences, but also take advantage of similarity/dissimilarity between objects in space. Four sets of hierarchical decision classes were tested based on 1) elevation derivatives only; 2) vegetation structural characteristics only; 3) spectral classification of land cover types; and 4) all variables, combined.

Following Zhu et al. (2006), we use spectral and structural data to create a multi-resolution image hierarchy frame to determine the most appropriate spatial resolution per land cover (object) type (Fig. 3). Up scaling is based on pixel similarity in space and includes scaling grids from 1 m to 30 m, using variable search radii (Franke, 1982) (Fig. 3). Up to 92 layers were assessed per land cover type prior to aggregation and removal of co-varying layers within classes used in the hierarchical classification. Probabilities are assigned to sub-layers relative to the training class, and these are accumulated to give an overall probability (weight) of prediction used to determine percent accuracy (and tested on study sub-areas, Fig. 1). Iterative (Boolean) decisions are made to keep or discard the sub-decision-tree in favour of other predictors based on a range of conditions used to estimate the spatial coverage of each land cover type (e.g. Bou Kheir, Böcher, Greve, & Greve, 2010) and confusion.

We also compare with a more typically used supervised classification routine applied to varying WorldView 2 spectral bands (shown as spectral signatures in Fig. 3 per land cover type). Several commonly used supervised and unsupervised classifications (including maximum likelihood and k-means clustering) were initially applied (e.g. Lillesand, Kiefer, & Chipman, 2003), but we found that the parallelepiped classification was most realistic, and will be compared for the remainder of the analysis. Best overall accuracies varied between 90% (kappa = 0.91) (parallelepiped); 81% (kappa = 0.70) (maximum likelihood); 76% (kappa = 0.61) (minimum distance); and 56% (kappa = 0.44) (k-means). The training sets used in the hierarchical classification were also used in the supervised classification.

3.4. Accuracy assessment and statistical analysis

Classification accuracies are identified using pixel confusion matrices per land cover type and per hierarchical class based on training sets and delineated plateau, bog, and fen sub-areas. Errors of omission and commission of a) the final (fusion) classification, and sensitivity

Fig. 2. Flow diagrams of initial LiDAR data processing used to create a digital elevation model (DEM), slope and aspect raster models, digital surface models (DSM), canopy height models (CHM), canopy gap fraction (gapfr) and intensity grids used as inputs to the DT classification. Bold font indicates an output, rounded boxes indicate an output used in the classification, and italics outline the method used.
Fig. 3. Flow diagrams of major steps used in the order of upland, permafrost plateau, bog, fen, and water land cover types. Bold text represents output data layers created and combined within the final classification based on decision criteria; grey boxes represent smoothing/filtering functions used to reduce “speckle” within classification stages; rounded boxes indicate that a previous output was used within a subsequent classification methodology, and italics represent an intermediate method used.
to b) topographic; c) canopy structure; and d) spectral sub-trees are presented as percent area confusion, and spatially within maps. The accuracy of edge delineation and land cover prediction using the hierarchical classification vs. the spectral classification are compared using the residual (shortest) distance between in situ plateau edge measurements and the classification method.

3.5. Application of a hydrological model

Several studies conducted at Scotty Creek have found spatial similarity in key terms that may be used as inputs to a simple hydrological model used to characterise hydraulic response of permafrost plateaus. These include: 1) rate of active layer thaw (Hayashi, Goeller, Quinton, & Wright, 2007); 2) hydraulic conductivity with depth below the ground surface (Quinton, Hayashi, & Carey, 2008), and 3) slope angles of plateau flanks separating relatively flat interiors and adjacent wetlands (Quinton & Baltzer, 2013; mean = 0.041, stdev. = 0.006).

Sub-surface discharge from plateaus is modelled using the distributed Cold Regions Hydrological Model (CHRM) (Pomeroy et al., 2007; Quinton & Baltzer, 2013). The purpose of the model is to simulate the hydrological cycle from hillslopes to medium-sized basins within a cold region context. The model requires the slope gradient, overall...
active layer thickness, initial position of the top of the frozen, saturated layer (Quinton & Hayashi, 2007), initial water equivalent of the snowpack (SWE), and the number of soil layers defined. Each soil layer is also characterised by thermal (e.g. volumetric heat capacity) and physical (e.g. bulk density, porosity) properties as well as the initial soil temperature as described in Quinton and Baltzer (2013). Additional input variables include soil air and ground surface temperatures, soil temperature and moisture content for each computational layer of soil, precipitation, water equivalent depth of a snowpack and the depth to the frost table measured at 30-minute intervals on a representative permafrost plateau. Air temperature was used to estimate the snowmelt rate, ground surface temperature, and ground thaw rate (Quinton & Gray, 2003) determined from trial and error until the computed SWE depletion matched observed depletion. These index methods ensure a close match between measured vs. simulated snowmelt and frost table recession thereby reducing the possibility that significant errors in runoff simulations are caused by snowmelt and ground thaw routines.

Computations were made at 30-minute time intervals for 15, 1-m wide strips of ground extending a) from the centre of the plateau across the relatively non-varying plateau top; and b) from the point of convexity where the flank of the plateau begins and including: i) the length to the edge of the hierarchical fusion classification; and ii) the length to the edge of the spectral (parallelepiped) classification from June 1st to August 31st, 2010. Transects also have 3D properties, where the upper boundary was located at the ground surface, and the lower boundary at the impermeable frost table (with depth into the soil). All inputs for individual transect runs were the same except for differences in slope and length of the plateau flank or sloped edge. Details on the model parameterisation, sensitivity and uncertainty analyses are provided in Quinton and Baltzer (2013).

4. Results

4.1. Classification results and comparisons

The accuracy of each DT classification methodology per land cover type is presented in Fig. 4, and spatially in Fig. 5. The hierarchical fusion of LiDAR data derivatives and spectra is the most accurate, explaining between 88% and 97% of the area extent of land cover classes within test areas and the broader watershed, whilst the spectral (parallelepiped) classification explained between 38% and 74% of the same area, following application of the same aggregation and de-speckling/aggregation methods. Within all land cover types, excluding uplands, topographic derivatives based on elevation are best able to classify land cover types, whilst vegetation structural characteristics and variable land cover spectra are generally less accurate. The parallelepiped classification of land covers has the greatest omission errors, but least confusion with other land cover types. Errors of omission are more important than commission as commission errors are subtracted out of the final land cover classification, in each case.

Between 82% and 96% of land cover extents and positions are identified from topographic derivatives, whilst only 69% of area coverage of uplands are identified using topography alone (vegetation structural characteristics are slightly better able to predict upland areas due to tall, dense vegetation unique to these areas). Vegetation structural characteristics (vegetation height, gap fraction) explain between 41%
and 76% of land cover area and location, with greatest correspondence (and least misclassified error) within uplands.

Errors of commission between other land cover types are greatest using vegetation structure as the hierarchical classifier due to confusion between typically short bog and fen vegetation in both (and methodologies, Fig. 3) and along the plateau edge where vegetation that was once located on thawed plateaus remains within the new land cover type (bog). The parallelepiped classification has greatest confusion within bogs due to heterogeneous spectral reflectance/absorption at the edges of plateaus and again tree encroachment into bogs (Fig. 5). The DT (fusion) classification also shows some confusion and misclassification within bogs and fens. Both have similar vegetation structural characteristics, are topographically low-lying within the landscape, and are spectrally similar in many parts due to same plant species types and soil moisture regimes. However, greater than 90% correspondence between bog and fen DT classification and the validation dataset is found within bog and fen areas, relative to 41% to 83% classification error using topographic (DEM), vegetation structural (CHM) and spectral methods.

The DT fusion classification combining all three hierarchical methods is illustrated in comparison with a more typically applied parallelepiped classification of high resolution WorldView 2 data (Fig. 6). Uplands have the greatest percent area coverage of the watershed (48%) followed by plateaus and bogs (20%, 19%), fens (12%) and lakes (2%). However, based on the spectral classification, plateaus cover the greatest percent area coverage of the watershed (43%, also found in Quinton et al. (2003) at Scotty Creek using an IKONOS spectral classification), followed by uplands (25%), fens (18%), bogs (12%), and water (3%). Differences range between 1% and 23% area coverage of land cover types within the watershed using the two methods. Plateaus have the largest
differences in extent between the two classification methods (23%), which will have significant implications for land surface and hydrological modelling, depending on which classification is used.

4.2. Validation and comparison of DT fusion classification and parallelepiped classification

Two final classification methods are compared with in situ validation transects of land cover type and plateau/fen edge delineation to quantify how well both classifications corresponded with measured, beyond indices of confusion or correspondence between pixels. Table 1 shows the percentage of pixels accurately identified as a certain land cover type compared with 245 geographically located measurements of environmental variables and visual observations along 10 transects in plateau, fen and bog land cover types. The DT fusion classification is better able to accurately predict land cover class within 8 of 10 transects, with accuracies ranging from between 78% and 100% and cumulative (average) land cover classification of 90%. The spectral classification successfully classifies land cover types between 55% and 100% of the time. Spectral classification along Transects 6 and 9 are an improvement over the DT classification method, whilst the cumulative (average) spectral classification success is 70%. The greatest differences between classification methods occur along Transects 4, 7, 8, and 10, where spectral characteristics between plateau and fen/bog land cover types are confounded by soil saturation (absorption in NIR bands) and shadowing along north-east margins. Differences may also have been exacerbated due to the late (October) data acquisition of the WorldView 2 dataset, and variable soil saturation conditions. Along Transect 6, the edge of the permafrost extent is underestimated using the DT fusion method because the plateau edge cannot be defined using the DEM, however this is the edge of the plateau and also an abundance of black spruce conifer trees, which are easily identified using the spectral classification (in correspondence with field measurements). The spectral classification corresponds better with ground measurements along Transect 9 because, mid-way through the transect, the fusion classification identifies a trail as “bog” (due to indentation and thaw into the plateau and reduction of tall vegetation representative of bogs within the classification).

The trail is not identified within the field measurements (classed as “plateau”) resulting in differences in accuracy between the methodologies.

The residuals between the geographic delineation of the water line between plateau and fen via in situ measurements (Fig. 1c) and nearest adjacent pixels of the DT fusion vs. spectral classifications are presented in Fig. 7. The fusion classification more accurately delineates the edge of the plateau, defined by the waterline, and is within 2 m of measured, 60% of the time, whereas the spectral classification achieves this 40% of the time, possibly as a result of confusion between shadows and saturated ground conditions that vary from time of WorldView 2 vs. in situ data collections. Two large misclassification errors in fusion classification, however, result in root mean squared error (RMSE) = 4.95 m (fusion), vs. 3.87 m (spectral) at distances of 139 m and 149 m along the transect. Differences in measured vs. classified are due to an area of subsidence within a permafrost plateau that was not measured using the GPS transect. The subsidence area was surrounded by trees at the outer edge, and was not easily visible when performing the water line survey (therefore, it was missed). Interestingly, the edge of the subsiding plateau is identified accurately by the fusion classification, when compared with the DEM, but is not identified by the spectral classification. If these two points are removed, RMSE is reduced to 2.10 m (DT fusion).

Residual differences between the spectral classification and the plateau edge/water line are also greater than 3 m for the first 7 measurement locations corresponding with vascular vegetation growth within the fen to distances of up to ~9 m beyond the transect (with greater errors corresponding to further extension of trees into the fen). Greatest residual differences between the DT fusion classification and the transect occur in the last 6 measurement locations (errors >3 m), corresponding to short, sparse vegetation along the southern margin of the plateau and therefore limitations to the decision criteria used to determine plateau extent.

4.3. CRHM model application and sensitivity analysis

The deviation in modelled total discharge (m$^3$) over 15 transects increases by 26% per unit area of the difference in the classification.
error, assuming that the DT fusion classification provides the most accurate estimate of plateau edge. Quinton and Baltzer (2013) found that a reduction in plateau runoff producing area had the most significant influence on plateau discharge, reducing it by almost half the volume between 2002 and 2010. This assumes that permafrost thaw results in reduced hydraulic gradient, increased thaw depth, and loss of runoff production. An overestimation in plateau area by 20% may artificially inflate discharge estimates by almost 1.3 m³ from June to August of a single year. Given the spectral classification almost doubles the area extent of permafrost plateaus compared with the DT fusion approach, the spectral classification would over-estimate permafrost thaw-related runoff from plateaus by a commensurate amount. This also suggests that anticipated plateau runoff from permafrost thaw may be lower than previously anticipated.

5. Discussion

Land cover type, location and extent directly influence processes related to hydrology and the parameterisation of hydrological models. These include (but are not limited to): land cover-specific energy balance, evapotranspiration, infiltration, and runoff (e.g. Chen, Chen, Ju, & Geng, 2005; Whitfield, St-Hilaire, & van der Kamp, 2009). Key spatial inputs to models often include spatial attributes, such as topography, geology, soils, land cover, and land use (Miller et al., 2007) for characterisation of land cover-specific processes. Many hydrological models used today are distributed or semi-distributed, requiring data specific to individual land cover types, but can also be simplified to general or empirical relationships (e.g. Govind et al., 2009; Kite, 1998). The response of hydrographs to precipitation/thaw from key land cover types within a distributed hydrological model may be organised in terms of a) Hydrological Response Units (HRU) based on land cover type, slope and aspect (e.g. Kite & Kouwen, 1992); b) irregular shapes (Kite & Pietroniro, 1996); or c) individual remote sensing pixels that contain unique vegetation–soil systems (excluding exchanges) (e.g. Zhang et al., 2012). It must be recognised, however, that whilst accurate classification of first order (mean) spatial variability is important, increasingly detailed characterisation of the land surface may introduce uncertainties as a result of location and extent errors in spatial datasets (Chen et al., 2005).

Whilst using remote sensing data has improved runoff predictions from distributed models (compared with lumped), some studies have found that small classification errors can result in substantial uncertainty to runoff and water quality (Miller et al., 2007). For example, Kite and Pietroniro (1996) found that differences in pixel resolution can have a significant influence on the distribution and extent of land cover types used within hydrological models. Further, Pietroniro, Prowse, Hamlin, Kouwen, and Soulsi (1996) found that results were improved when calibrating a hydrological model to an accurate land cover classification (determined from a confusion matrix), however, the model was still not able to adequately reproduce the volume and timing of runoff. They attributed model inaccuracies to the simplification of land cover processes, and a lack of detailed information on storage and routing processes related to land cover type.

In the case of the Scotty Creek watershed, we have found that a few studies have performed spectral and textural classifications using remote sensing datasets for the purposes of hydrological modelling and/or improved understanding of runoff processes. In all cases, the classifications of land cover extents were quite different. Pietroniro et al. (1996) used a principal components analysis and 8 spectral bands from a Landsat image within a k-means unsupervised classification. They found that wetland areas covered ~22% of three study basins adjacent to and including Scotty Creek, whilst forests covered an additional ~73% of the total basin area (subdivided by vegetation type: deciduous, coniferous, mixed and transitional). A study by Quinton et al. (2003), who used Landsat and IKONOS remote sensing data classified using a maximum likelihood classification, found that the Scotty Creek watershed was comprised of 63% “wooded” areas (representing uplands and permafrost plateaus), 19.6% fen, and 10.2% bog. In another study by Stadnyk et al. (2005), they conclude that the classification of bogs and fens within the Scotty Creek watershed using Landsat data was significantly underestimated. They characterised the distribution of wetlands to cover 13% of the watershed, with forested areas, (conifer, mixed, and transitional) covering an additional 87% of the watershed.

The results presented in our study indicate that bogs and fens represent 19% and 12% of the watershed, respectively, which indicate possible confusion between land cover types/pixels, and mixed pixel small bog omissions using lower resolution remote sensing data. Fen areas may be less prone to edge uncertainties because they typically cover much broader areas and are less prone to micro-scale variability at fen edges (unlike bogs, which are often fragmented and may be in partial shadow which may be confused with plateau land cover types, e.g. Chasmer et al., 2010). Uplands and plateaus accounted for 63% of the watershed in Quinton et al. (2003), which is similar to the finding in this study (68%). Because the magnitude and timing of runoff processes are strongly related to water routing and retention capacity of bogs and fens, accurate classification of location, extent and connectivity of wetland areas in the Scotty Creek watershed is very important. Examples of the influence of wetlands to basin runoff are demonstrated in Quinton et al. (2003) and Stadnyk et al. (2005). When compared with other basins within the Lower Liard, Quinton et al. (2003) found that basins with greater area coverage of fens had greater annual basin runoff than those with lower proportional coverage of fens to bogs. Both Stadnyk et al. (2005) and Quinton et al. (2003) note the importance of hydrologically connected wetlands as conveyers of water to the basin outlet, whilst Stadnyk et al. (2005) state that accurate classification of these connected areas as a function of topography is a critical input required for accurate prediction of basin runoff.

Clearly, the dependence on classification accuracies from a confusion matrix with kappa values must be regarded with some caution, as indicated in this study and others. Classification accuracies are used to statistically assess the representation of land covers by spectra (e.g. Foody, 2002), but may not entirely or always represent reality. In this study, the best parallelepiped classification of the Scotty Creek watershed generated average accuracies of ~88%, with an overall accuracy of ~91% (p < 0.001, kappa = 0.91) (accuracies ranged between ~78% and 98% based on training sets, errors of omission and commission within a confusion matrix). However, when compared with validation data, the classification was accurate between 38% and 74% of the time. Pietroniro et al. (1996) present similar accuracies for a Landsat classification of three watersheds, including Scotty Creek (91% overall accuracy, kappa = 0.91), whilst Quinton et al. (2003) and Stadnyk et al. (2005) do not quantify classification errors. From these results we recommend caution when stating classification accuracies from confusion matrices, but understand that in many cases, the confusion matrix is a best estimate of the error associated with spectral misclassification.

---

### Table 1

Percent of land pixels correctly classified using the hierarchical fusion classification introduced in this study vs. a parallelepiped spectral classification per land cover type compared with geographically referenced field assessments.

<table>
<thead>
<tr>
<th></th>
<th>T1 (n = 18)</th>
<th>T2 (n = 20)</th>
<th>T3 (n = 21)</th>
<th>T4 (n = 31)</th>
<th>T5 (n = 25)</th>
<th>T6 (n = 18)</th>
<th>T7 (n = 20)</th>
<th>T8 (n = 31)</th>
<th>T9 (n = 18)</th>
<th>T10 (n = 43)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hierarchical fusion classification</strong></td>
<td>78</td>
<td>91</td>
<td>100</td>
<td>97</td>
<td>96</td>
<td>89</td>
<td>85</td>
<td>94</td>
<td>94</td>
<td>84</td>
</tr>
<tr>
<td><strong>Spectral classification</strong></td>
<td>61</td>
<td>85</td>
<td>85</td>
<td>75</td>
<td>88</td>
<td>94</td>
<td>50</td>
<td>61</td>
<td>100</td>
<td>55</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>17</td>
<td>6</td>
<td>15</td>
<td>23</td>
<td>8</td>
<td>–6</td>
<td>35</td>
<td>32</td>
<td>–6</td>
<td>29</td>
</tr>
</tbody>
</table>
6. Conclusions

In this study, we combine topographic derivatives with vegetation structural and spectral characteristics unique to land cover types found at the edge of the southerly margin of discontinuous (sporadic) permafrost regions of northern Canada. The methodology developed here significantly advances current spectral classification methodologies used to identify areas of permafrost plateaus, surrounded by saturated fens, and pocked by connected/isolated bogs. Further, this is a requirement both for accurate hydrological modelling and improved understanding of water runoff processes at the southern-most margin of discontinuous permafrost (Quinton et al., 2003; Stadnyk et al., 2005). Although both spectral and vegetation structural characteristics improved the classification slightly, the use of topographic derivatives provided the greatest explanation of land cover variability, especially where low-lying areas and channels could be identified (e.g. Stadnyk et al., 2005). Therefore, high resolution spectral data are not necessarily required for input into the hierarchical classification. In comparison, results from supervised (and unsupervised) spectral classifications in this area should be regarded with some caution in this region. We find that the results of the correlation matrix do not necessarily match validation data. Alternative methods for assessing classification errors within the zone of discontinuous permafrost should include another form of validation (e.g. transects) or possibly training from an accurate classification like the one presented in this study.

Acknowledgements

Funding for this research was provided by the Natural Sciences and Engineering Research Council, and the Canadian Space Agency. We are grateful to Allyson Fox for help with the LiDAR data acquisition, Dr. Tristan Goulden, and Alex MacLean for GPS surveys. We would also like to thank Kellina Higgins for ground cover vegetation surveys and analysis used in this study, and Dr. Oliver Sonnentag for helpful conversations regarding vegetation classification and accuracy. We are thankful for the helpful comments from an anonymous reviewer. Research was supported by the Dehcho First Nations, Fort Simpson and Kiiilii Kue First Nation.

References


