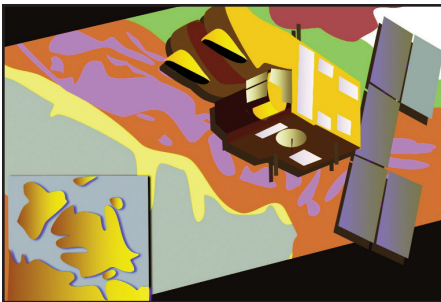


SERIES



Remote Predictive Mapping 6. A Comparison of Different Remotely Sensed Data for Classifying Bedrock Types in Canada's Arctic: Application of the Robust Classification Method and Random Forests

Harris, J.R.¹, He Xia Juan², Rainbird, R.¹, and Behnia, P.¹

¹*Geological Survey of Canada
601 Booth St.
Ottawa, Ontario, K1A 0E8, Canada
E-mail: harris@nrcan.gc.ca*

²*Faculty of Geography
Ottawa University
Simard Hall, 60 University St.
Ottawa, Ontario, K1N 6N5, Canada*

SUMMARY

The Geological Survey of Canada, under the Remote Predictive Mapping project of the Geo-mapping for Energy and Minerals program, Natural

Resources Canada, has the mandate to produce up-to-date geoscience maps of Canada's territory north of latitude 60°. Over the past three decades, the increased availability of space-borne sensors imaging the Earth's surface using increasingly higher spatial and spectral resolutions has allowed geologic remote sensing to evolve from being primarily a qualitative discipline to a quantitative discipline based on the computer analysis of digital images.

Classification of remotely sensed data is a well-known and common image processing application that has been used since the early 1970s, concomitant with the launch of the first Landsat (ERTS) earth observational satellite. In this study, supervised classification is employed using a new algorithm known as the Robust Classification Method (RCM), as well as a Random Forest (RF) classifier, to a variety of remotely sensed data including Landsat-7, Landsat-8, Spot-5, Aster and airborne magnetic imagery, producing predictions (classifications) of bedrock lithology and Quaternary cover in central Victoria Island, Northwest Territories. The different data types are compared and contrasted to evaluate how well they classify various lithotypes and surficial materials; these evaluations are validated by confusion analysis (confusion matrices) as well as by comparing the generalized classifications with the newly produced geology map of the study area. In addition, three new Multiple Classification System (MCS) methods are proposed that leverage the best characteristics of all remotely sensed data used for classification.

Both RCM (using the maximum likelihood classification algorithm, or MLC) and RF provide good classification results; however, RF provides the highest classification accuracy because it uses all 43 of the raw and derived bands from all remotely sensed data. The MCS classifications, based on the generalized training dataset, show the best agreement with the new geology map for the study area.

SOMMAIRE

Dans le cadre de son projet de Télécartographie prédictive du Programme de géocartographie de l'énergie et des minéraux de Ressources naturelles Canada, la Commission géologique du Canada a le mandat de produire des cartes géoscientifiques à jour du territoire du Canada au nord de la latitude 60°. Au cours des trois dernières décennies, le nombre croissant des détecteurs aérospatiaux aux résolutions spatiales et spectrales de plus en plus élevées a fait passer la télédétection géologique d'une discipline principalement qualitative à une discipline quantitative basée sur l'analyse informatique d'images numériques.

La classification des données de télédétection est une application commune et bien connue de traitement d'image qui est utilisée depuis le début des années 1970, parallèlement au lancement de Landsat (ERST) le premier satellite d'observation de la Terre. Dans le cas présent, nous avons employé une méthode de classification dirigée en ayant recours à un nouvel algorithme appelé Méthode de classification robuste (MRC), ainsi qu'au classificateur Random Forest (RF), appliqués à une variété de données de

téledétection dont celles de Landsat-7, Landsat-8, Spot-5, Aster et d'imagerie magnétique aéroportée, pour produire des classifications prédictives de la lithologie du substratum rocheux et de la couverture Quaternaire du centre de l'île Victoria, dans les Territoires du Nord-Ouest. Les différents types de données sont comparés et contrastés pour évaluer dans quelle mesure ils classent les divers lithotypes et matériaux de surface; ces évaluations sont validés par analyse de matrices de confusion et par comparaison des classifications généralisées des nouvelles cartes géologiques de la zone d'étude. En outre, trois nouvelles méthodes par système de classification multiple (MCS) sont proposées qui permettent d'exploiter les meilleures caractéristiques de toutes les données de télédétection utilisées pour la classification.

Tant la méthode MRC (utilisant l'algorithme de classification de vraisemblance maximale ou MLC que la méthode RF donne de bons résultats de classification; toutefois c'est la méthode RF qui offre la précision de classification la plus élevée car elle utilise toutes les 43 les bandes de données brutes et dérivées de toutes les données de télédétection. Les classifications MCS, basées sur le jeu de données généralisées d'apprentissage, montrent le meilleur accord avec la nouvelle carte géologique de la zone d'étude.

INTRODUCTION

Over the past three decades, the increased availability of space-borne sensors imaging the Earth's surface using increasingly higher spatial and spectral resolutions has allowed geologic remote sensing to evolve from being primarily a qualitative discipline to a quantitative discipline based on the computer analysis of digital images. Many image processing and analytical techniques have been developed to aid the interpretation and extraction of useful geologic information from remotely sensed imagery. Image classification is perhaps the best known and the most widely used digital image analysis technique since the advent of the ERTS-1 (Landsat) series of satellites in the early 1970s.

The Geological Survey of Canada, under the Remote Predictive Mapping (RPM) project (Desnoyers

and Harris 2003; Schetselaar et al. 2007; Harris 2008; Harris et al. 2008a), part of the Geo-mapping for Energy and Minerals (GEM) program of Natural Resources Canada, has the mandate to produce up-to-date geoscience maps of Canada's territory north of latitude 60°. The expansive territory to be mapped, in combination with the required resources and cost of setting up field mapping campaigns, requires new approaches to providing geological information. Remote Predictive Mapping (RPM), which emphasizes the use of various types of satellite and geophysical imagery to assist in producing geological maps, in part addresses this requirement (Harris 2008, and papers therein). Producing the required geological information can be undertaken by visual interpretation of enhanced imagery, by employing machine learning algorithms to assist in extracting information, or by a combination of the two methods (Schetselaar et al. 2007; Harris et al. 2008a, 2012a). Predictive mapping is obviously not a replacement for field work; however, the intelligent use of remotely sensed, geophysical and geochemical data in concert with legacy geological data (maps, databases) can produce maps that focus field work on areas that have more complex signatures (Harris 2008). Furthermore, RPM also provides first-order geologic predictive maps in poorly mapped areas or areas where field work is not possible.

The use of machine learning algorithms, primarily classification, for mapping rocks has been applied mostly in arid environments (Rowan et al. 1987; Macias 1995; Glikson and Creasey 1995; El Rakaiby 1995; Van de Meer et al. 1995; Rowan et al. 2005; Kavak 2005; Peña and Abdelsalam 2006; Saadi and Watanabe 2009; Rajendran et al. 2012). A number of examples can also be found for Arctic environments (Leverington 2001, 2010; Lorenz 2004; Harris et al. 2005, 2008a, c, 2009, 2010; Wickert et al. 2008; Schetselaar and Ryan 2008; Leverington and Moon 2012; Behnia et al. 2012). Even though Arctic environments are not covered by trees, ubiquitous lichen and snow cover and a short season for acquisition of suitable imagery can make the application of machine learning algorithms for geo-

logical mapping a challenge. However, the studies listed above for northern environments do indicate that the classification of various remotely sensed data, including geophysical data, can make a useful contribution to mapping activities. Copious public domain remotely sensed data (www.geobase.ca) are freely available and can be applied to operational mapping programs such as RPM using both visual interpretation and machine learning methods. The application of machine learning algorithms offers the advantages of objectivity, speed and the ability to cover large areas efficiently and within short time frames. However, expert knowledge is paramount to the success of machine learning algorithms, as training data are required for the geological features to be successfully classified.

Classification algorithms can be broadly divided into two categories – unsupervised and supervised. Unsupervised algorithms require very little user input and are designed to find like classes (or clusters) within an *n*-dimensional data (feature) space, whereas supervised classification requires much more user input to train and guide the classification process. Supervised classification is based on *a priori* knowledge of the features to be classified. The spectral signature of representative samples of the different surface cover types (information classes), referred to as *training areas*, are used to train the classifier and, by using a classification algorithm, to recognize spectrally similar areas for each class on an image dataset.

In this study, supervised classification using a new algorithm known as the Robust Classification Method (Harris et al. 2012b) is applied to a variety of remotely sensed data, including Landsat-7, Landsat-8, Spot-5, Aster and airborne magnetic imagery, producing predictions (classifications) of bedrock lithology and Quaternary cover in central Victoria Island, Northwest Territories. A Random Forest classifier is also employed to (1) rank the predictive power of each input band, and (2) create a classified map. Both of these classifiers are 'ensemble classifiers', as they produce multiple predictions (classified maps) that are combined using a voting procedure

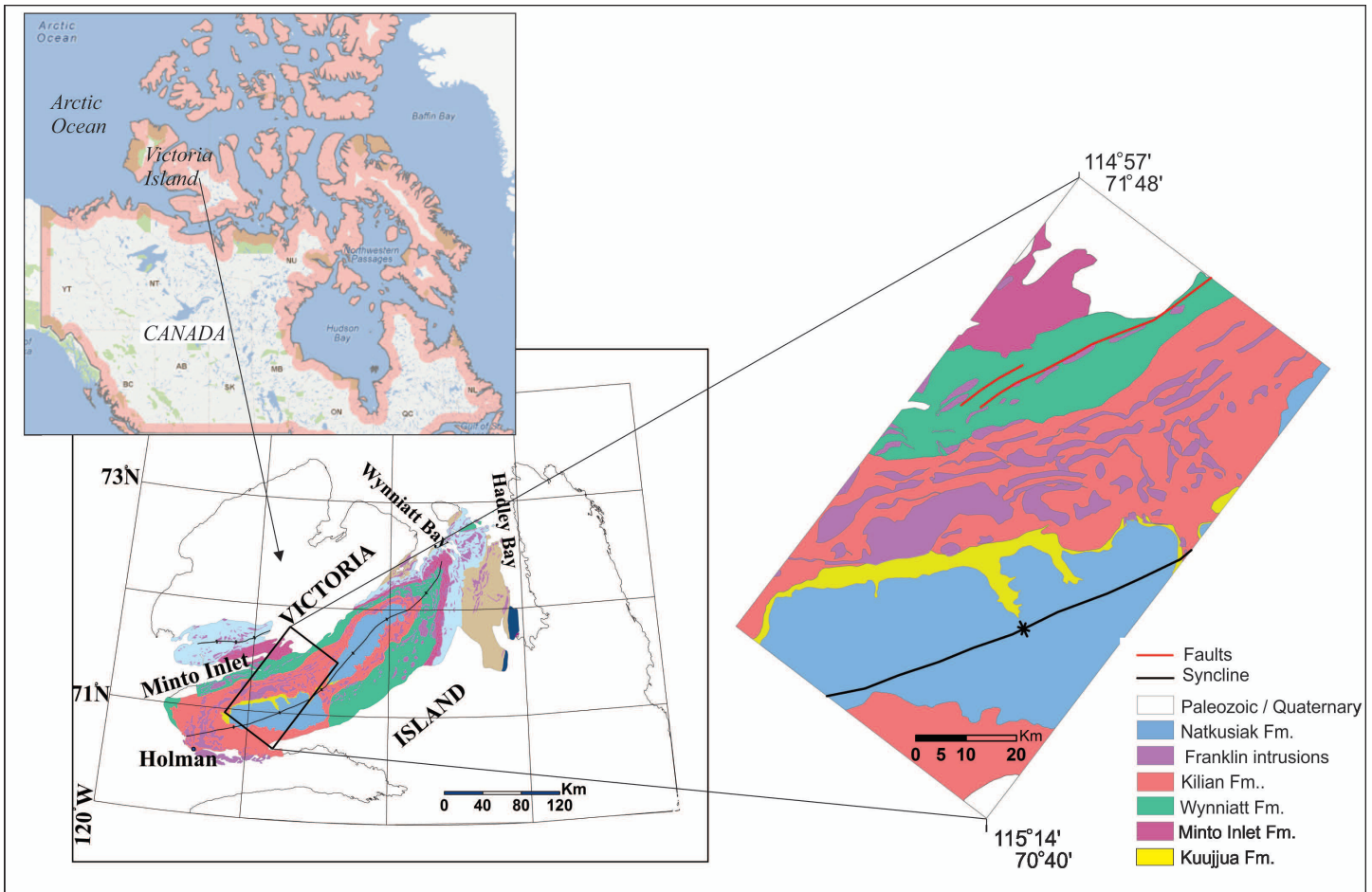


Figure 1. Location of study area in the Minto Inlier of Victoria Island, Northwest Territories, Canada, and legacy geology map showing main lithological units (geology from Hubert et al. (2005), modified from Thorsteinsson and Tozer (1962).

whereby the predominant class, on a pixel basis, is selected for the final map. Using confusion analysis (confusion matrices) and comparing the generalized classifications with the newly produced geology map of the study area (Rainbird et al. 2013a, b, c and d), these different data types are compared and contrasted to evaluate how well they classify various lithotypes and surficial materials. We also introduce three new multiple classification system (MCS) maps that compare favourably to the geology map of the study area. ENVI™ image processing software in concert with ArcGIS™ and the EnMap Box software package (<http://www.enmap.org/?q=enmapbox>) were used to perform the classifications.

STUDY AREA

The study area lies within the Minto Inlier on Victoria Island (Fig. 1). The Minto Inlier, first recognized and mapped by Thorsteinsson and Tozer

(1962), is a northeast-trending belt of early Neoproterozoic sedimentary rocks of the Shaler Supergroup that are intruded by gabbro–diorite sills and dykes and capped by coeval flood basalt of the Franklin intrusive (magmatic) event. The succession was gently folded, forming an open syncline and a smaller anticline, before deposition of an unconformably overlying shallow marine sandstone and carbonate succession of Cambro–Ordovician age. The Shaler Supergroup includes, in ascending stratigraphic order, the Rae Group, Reynolds Point Group, Minto Inlet Formation, Wynniatt Formation, Kilian Formation and Kuujjua Formation (Rainbird et al. 1994, 1996; Fig. 1). With the aid of recent detailed field observations, the Wynniatt Formation was divided into four lithostratigraphic sub-units, namely black shale, lower carbonate, stromatolitic carbonate, and upper carbonate members (Thomson et al. 2014). Basaltic

rocks of the Natkusiak Formation are exposed in the south, and related gabbro–diorite sills and dykes (Franklin Intrusions) were intruded into the sedimentary strata throughout the area. Recently, new geological maps of the study area have been produced using detailed field observations and visual interpretation of Spot-5 stereo imagery (Rainbird et al. 2013a, b, c and d). A portion of these maps is presented in Figure 2a, which shows the detailed geology, whereas Figure 2b is a generalized lithological map that will be used for comparisons with the generalized lithological and surficial classification maps produced in this study. Table 1 includes descriptions of the geology shown in Figure 2a (detailed lithology and surficial materials column), as well as the generalized lithology in Figure 2b (generalized lithology and surficial materials column).

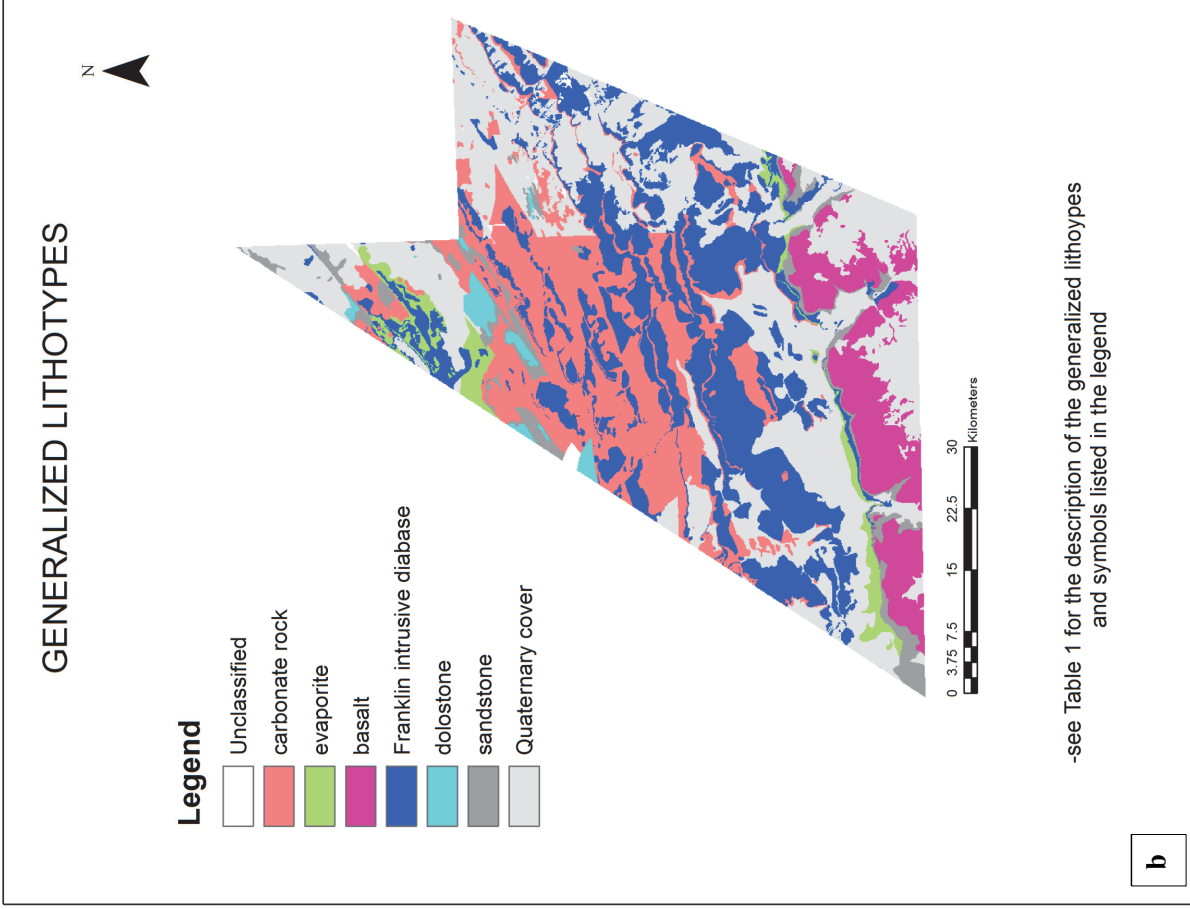
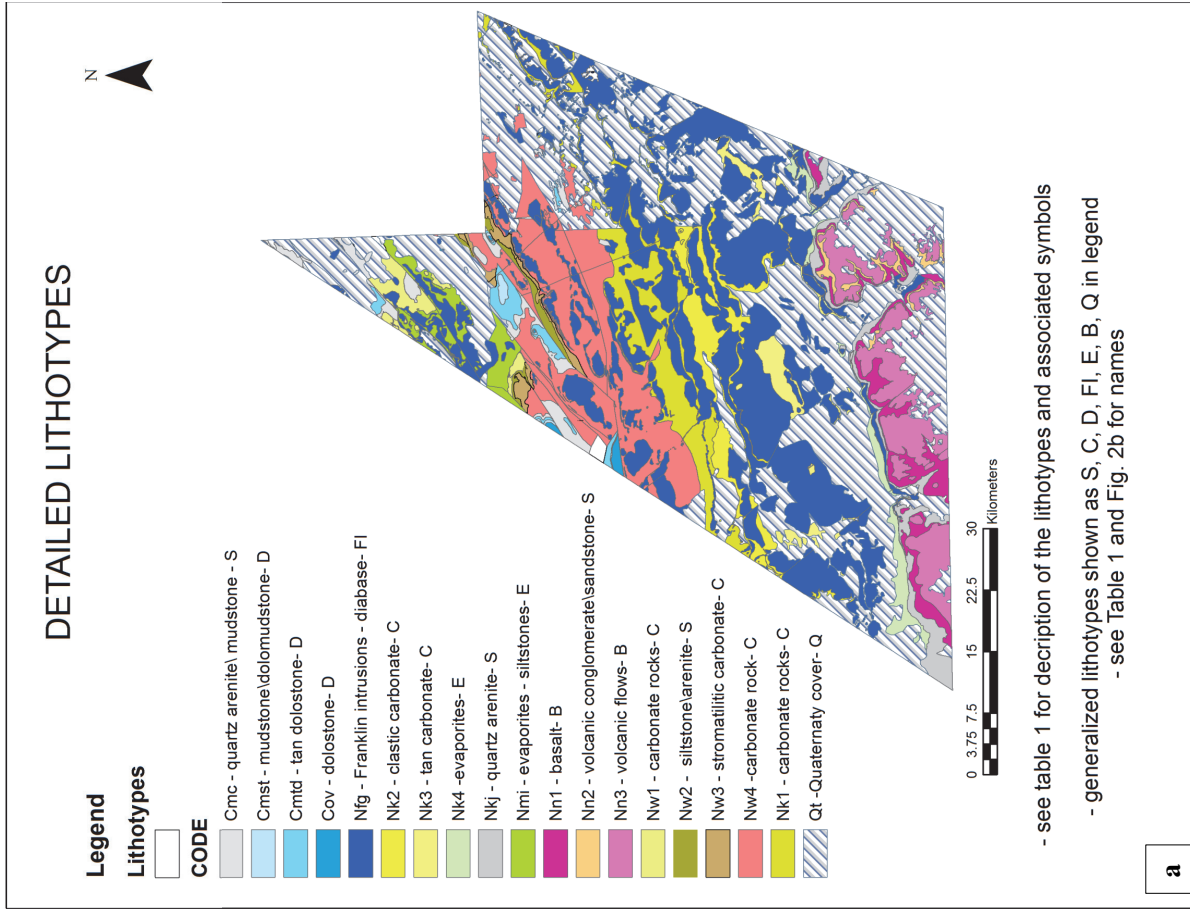


Figure 2. New geology map of the study area in the Minto Inlier of Victoria Island, Northwest Territories, Canada. (a) Detailed bedrock geology; see Table 1 for descriptions of lithological units. (b) Generalized lithology; see Table 1 for a description of how the generalized lithotypes were derived from Figure 2a. Geology modified from Rainbird et al. (2013a, b, c, d).

Table 1. Description of bedrock types and surficial classes (training areas) used for classification of the Landsat, Spot and Aster data using Robust Classification Method–maximum likelihood classification (RCM–MLC) and Random Forest (RF). A total of thirteen lithological – surficial types were classified. These thirteen categories were generalized into seven major lithotypes, based on the detailed lithology (Figure 2a) for comparison with the new generalized geology map (Figure 2b).

| Detailed lithologies and surficial material | Description | Training area | Generalized lithologies and surficial materials |
|--|--|---|---|
| (Fig. 2a – see legend – symbols listed below) | | (13 spectral units used for classification) | (Fig.2b – 7 generalized lithological-surficial classes) |
| Natkusiak basalt (1) –Nn1 | Extends SE-NW in southern part. Low reflectance in VNIR. Moderate reflectance in SWIR. | Nfb1- (4 polygons – 184 pixels) | Basalt |
| Natkusiak basalt (2) – Nn2 | Low reflectance in VNIR and SWIR. Distinguishable from basalt (1) in SWIR bands giving a layering appearance. | Nfb2 – (4 polygons – 179 pixels) | |
| Natkusiak basalt (3) –Nn3 | Basalts covered by thin vegetation layer. | Nfbv – (5 polygons – 87 pixels) | |
| Franklin Intrusives =- diabase (1) -Nfg | Extends in the central and northern parts of the area. Very similar to basalts in VNIR. Has slightly higher reflectance in SWIR-2 compared to basalts. | Nfg – (9 polygons – 420 pixels) | Franklin intrusive rocks (diabase) |
| Franklin intrusives - diabase (2)-Nfg | Gabbro/diorite covered by thin vegetation | Nfgv – (4 polygons – 187 pixels) | |
| Wynniatt Fm. (carbonates) Nw1, , Nw3, Nw4 | Include all carbonate members of Wynniatt Fm. High reflectance in SWIR, moderate in VNIR. | Carbonates – (10 polygons - 293 pixels) | Carbonates |
| Cambro/Ordovician (carbonates)- Nk1, Nk2, Nk3,Nk4 | Very similar spectral response with Wynniatt FM. Slightly different in SWIR bands. | | |
| Paleozoic dolostones- CmtD, Cov, Cmst | high spectral response with | Dolostones – 6 polygons – 82 pixels) | Dolostones |
| Kuujua Fm.(quartz arenites) – Njk Nw2, Cmc – clastic unit – mostly covered by vegetation | Occur in contact with the basalts. Low to moderate reflection in VNIR. Moderate reflectance in SWIR-1, moderate to high in SWIR-2. | Sandstones – (4 polygons – 178 pixels) | Sandstones |
| Evaporite rocks- Nmi (Minto Inlet Fm), Nk4 | Includes evaporates of Minto and Killian Fms. Low reflection in NIR, slightly higher reflection in SWIR. | Evaporites – (8 polygons – 140 pixels) | Evaporites |
| Till 1 -Q | | Till 1 – (6 polygons- 191 pixels) | Quaternary |
| Till 2-Q | | Till 2 – (8 polygons – 156 pixels) | |
| Till 3-Q | | Till 3 – (8 polygons – 224 pixels) | |
| Dry Veg | Dry veg missed by biomass mask | Dry veg | Not classified |

Table 2. Summary of remotely sensed data and associated acquisition characteristics.[^]

| | Number of bands | Spectral range | Spatial resolution | Dynamic range | Date | Sun Azimuth/ Elevation (degrees) |
|---|-----------------|---------------------------------------|--------------------|---|-----------|--|
| LANDSAT-7 | 3-V | 0.450–0.515 μm (B) | 30 m (XS) | 8 bits | 7/14/2002 | 172.9 / 40.1 |
| | 1-NIR | 0.525–0.605 μm (G) | 60 m (TIR) | | | |
| | 2-SWIR | 0.630–0.690 μm (R) | 15m (pan) | | | |
| | 2-TIR | 0.750–0.900 μm (NIR) | | | | |
| | 1-pan | 1.55–1.75 μm (SWIR) | | | | |
| | | 2.08–2.35 μm (SWIR) | | | | |
| 10.40–12.50 μm (TIR) | | | | | | |
| | | 0.520–0.900 μm (pan) | | | | |
| SPOT-5 | 2-V | 0.500-0.590 μm (G) | 10 m (XS) | 8 bits | 8/22/2009 | 174.3 / 30.4 |
| | 1-NIR | 0.610–0.680 μm (R) | 20m (SWI) | | | |
| | 1-SWIR | 0.780–0.890 μm (NIR) | 2.5m (pan) | | | |
| | 1-pan | 1.58–1.75 μm (SWIR) | | | | |
| | | 0.480–0.710 μm (pan) | | | | |
| LANDSAT - 8 | 4-V | 0.43-.45 (B) | 15 m (pan) | 16 bit | 7/4/2013 | 176.1 / 41.9 |
| | 1-NIR | 0.45 -0.51 (B) | 30m (XS) | | | |
| | 2-SWIR | 0.53 – 0.59 (G) | 30m (TIR) | | | |
| | 1-TIR | 0.64 – 0.67 (R) | | | | |
| | 1-pan | 0.85 – 0.88 (NIR) | | | | |
| | | 1.57 – 1.65 (SWIR) | | | | |
| | | 2.11 – 2.29 (SWIR) | | | | |
| | | 10.6-11.1 (TIR) | | | | |
| 11.5 -12.51 (TIR) | | | | | | |
| | | 0.5 – 0.68 (PAN) | | | | |
| ASTER (two mosaicked scenes) – 1B data (radiometric and geometric coefficients applied – radiance at sensor) | 3 –VNIR | 0.52-0.60 (G) | VNIR - 15m | Bands 1– 9 (VNIR and SWIR - 8 bit) Bands 10 –14 (TIR --16 bit) | 9/17/2005 | South scene – 181.1 / 20.7 North Scene – 182..3 / 20.2 |
| | 6 –SWIR | 0.63-0.69 (R) | SWIR – 30m | | | |
| | 6-TIR | 0.76 – 0.86 (NIR) | TIR – 90m | | | |
| | | 1.6 – 1.7 (SWIR) | | | | |
| | | 2.145 – 2.185 (SWIR) | | | | |
| | | 2.185 – 2.225 (SWIR) | | | | |
| | | 2.235 – 2.285 (SWIR) | | | | |
| | | 2.295 – 2.365 (SWIR) | | | | |
| | | 2.360 – 2.430 (SWIR) | | | | |
| | | 8.125 – 8.475 (TIR) | | | | |
| 8.475 – 8.825 (TIR) | | | | | | |
| 8.925 – 9.275 (TIR) | | | | | | |
| 10.25 – 10.95 (TIR) | | | | | | |
| 10.95 – 11.65 (TIR) | | | | | | |
| Airborne Magnetic Data | 1 channel | Total field (magnetic susceptibility) | | 16 bit | | 1 km line spacing and 60m sample spacing; gridded using a minimum curvature function |

[^] XS = multispectral, TIR = thermal, NIR = near infrared, SWIR = short-wave infrared, V = visible, μm = microns ($1* 10^6\text{m}$).

DATA AND PROCESSING

The types and descriptions of remotely sensed data used for classification purposes are presented in Table 2. Landsat-7 orthorectified images (both images acquired on 14 July, 2002) were downloaded from the United States Geological Survey (USGS) Global Visualization Viewer (<http://glovis.usgs.gov>), and two-level 1A Spot-5 scenes acquired on 22 August, 2009 were purchased from Blackbridge (<http://www.blackbridge.com/>). Both the Landsat-7 and Spot-5 data in 8-bit format were radiometrically balanced, mosaicked and then subset to the extent of the study area (Fig. 1). The Landsat-8 data acquired on 4 July, 2013 were downloaded from the USGS in 16-bit format; these data covered the entire study area. Two Aster scenes (LEVEL 1B) were purchased from the USGS; both were acquired on 17 September, 2005 at relatively low sun angles (Table 2). These were the best (cloud free) Aster images that could be sourced for the study area. However, the late season acquisition and consequent low sun angle may affect the classification results in a negative fashion (see Discussion section). The two images were radiometrically balanced and mosaicked.

Airborne total field magnetic data acquired by the Geological Survey of Canada (<http://gdr.aggrnrcan.gc.ca/gdrdap/dap/search-eng.php>) were also used to help classify the various rock types. These data were collected along north-south flight lines spaced 1 km apart. Sample spacing was approximately 60 m along-line and the data were gridded using a minimum curvature algorithm. The patterns of tones depicted on a total field magnetic map show the strength of the measured magnetic field and hence reflect how magnetic the rocks are. Trends in the anomalies therefore represent the distribution of magnetic minerals and reflect the geological make-up of the area.

All optical imagery was first corrected for atmospheric effects using the dark subtraction method (Gupta 1991) and enhanced by contrast stretching. In addition, the Aster radiance data were corrected for cross-talk, which is caused by signal leakage from band 4 into adjacent bands 5 and 9

(Kalinowski and Oliver 2004; Biggar et al. 2005). All images were visually and statistically assessed after all corrections were applied to ensure good quality (radiometrically balanced) data suitable for digital classification.

To obtain better discrimination between various lithological features, areas covered with water, ice and vegetation were masked using the Landsat-7 and Spot-5 imagery. The Landsat-7 image was used for masking vegetation as it was acquired in July when biomass would be most extensive. The well-known Normalized Difference Vegetation Index (NDVI) was used to characterize the vegetation distribution on the Landsat-7. NDVI, which has a dynamic range from -1 to +1 (Chuvieco and Huete 2010), is defined as:

$$NDVI = (NIR - RED) / (NIR + RED) \quad (Eq. 1)$$

where NIR = near infrared band and RED = red band.

An upper threshold was defined to separate vegetation from non-vegetation and was used to mask all imagery. Applying the Landsat-7 vegetation mask to all data was undertaken to facilitate an unbiased comparison between the different classifications. The Spot-5 data appeared to show the most snow and ice, so these data were used to create a snow and ice mask that was, once again, applied to all imagery. This was accomplished by applying an upper threshold to the histograms for Spot-5 bands 2 and 3 (green and red), in which snow and ice is highly reflective.

Because topography can affect spectral analysis (Drury 1993; Jensen 2005), a 1:50,000 Digital Elevation Model (DEM) mosaic of the Canadian Digital Elevation Data (downloaded from Canadian Council on Geomatics website: www.geobase.ca) was created for the study area, and shadows were modelled to produce hill shade images using the sun azimuth and elevation values for each sensor (see Table 2 for sun azimuth and elevation values). A mask was created from these images and overlaid on the final classification maps to screen out areas of topographic shadows and diffuse reflection from back slopes that could result in spectral misclassification.

Calculation of Ratios

Calculation of band ratios is a common image processing technique used to highlight spectral reflectance differences between minerals as a function of wavelength as well as to reduce topographic illumination effects (Drury 1993; Jensen 2005). Table 3 presents well-documented ratios applied to the optical data in this study in order to highlight different minerals. These ratios were also used in concert with the raw data to classify different rock types.

Training Areas

Once the images were masked, the spectral – lithological classes were defined (training area and description columns, Table 1). This was based on expert geologic knowledge of the study area as well as the spectral diversity observable in the Landsat, Spot and Aster images. Six main lithological classes (Table 1), including basalt of the Natkusiak Formation, Franklin intrusive rocks (gabbro–diorite sills and dykes), carbonate of the Wynniatt Formation, carbonate of the Cambro–Ordovician succession, evaporite rocks of the Minto Inlet and Killian Formations, and quartz arenite of the Kuu-jjua Formation (Fig. 1) were defined based on the new geological map (Fig. 2a) and field information.

Although vegetation, water and ice were excluded in all images by masking (discussed above), there were still some small areas of dry (low chlorophyll) vegetation present in the resulting images. To avoid the interference of these spectra with other rock spectra, an additional class was created to represent dry vegetation. Some of the classes that were selected on the basis of expert geologic knowledge showed variable spectral responses in different parts of the images. This could be attributed, in part, to the complicating spectral signatures of overlying glacial sediments and/or other overburden materials or vegetation cover. For example, in some parts of the study area, the Natkusiak basalt exhibits a lower reflectance that is visually distinguishable from neighbouring basaltic areas. This apparent spectral difference between basalts could be attributed to primary layering and compositional differences, grain size, and

Table 3. Summary of ratios applied to the remotely sensed data.[^]

| Data | Ratio (band numbers) | Comment |
|---------------------|--------------------------|--------------------------------|
| LANDSAT 7 | 3/1 | Ferric iron |
| | 4/3 | Biomass - vegetation |
| | 5/7 | Clay |
| | 5/4 | Ferrous iron |
| LANDSAT 8 | 6/5 | Ferrous iron |
| | 6/7 | Clay |
| | 5/4 | Biomass - vegetation |
| | 4/1b | Ferric iron |
| ASTER – VNIR & SWIR | 6/7 | Phyllic alteration |
| | 5/6 | Kaolinite, muscovite, phengite |
| | 5+7/6 | Phyllic alteration |
| | 6+9/8 | Amphibole / MgOH |
| | 6+9/7+8 | Epidote, chlorite, amphibole |
| | 7+9/8 | Epidote, chlorite, carbonate |
| | 4/3 | Ferric oxides |
| | 4/2 | Gossan |
| | 3/2 | Biomass - vegetation |
| | 4/1 | Gossan |
| | 2/1 | Ferric iron |
| 6/8 | Amphibole | |

[^]References for these ratios can be found in Drury (1993), Abrams and Hook (1995), Lillesand and Keiffer (2004), and Kanlinowski and Oliver (2004).

secondary processes such as weathering. To account for these apparent spectral differences, two separate basalt classes were assigned, as well as a third class for those parts that were covered by thin, dry vegetation. Similarly, a separate class was assigned to gabbro–diorite covered by thin vegetation. Three surficial classes (glacial tills) were defined based on the spectral responses observable on the Landsat-7, Landsat-8, and Spot-5 imagery, making a total of thirteen classes used for the classification (see Table 1).

To facilitate a comparison with the new geology map, both the detailed lithological map (Fig. 2a) and training areas were generalized into six basic rock types and one surficial type (Fig. 2b) to evaluate which type of imagery provided the closest representation of the mapped geology (Table 1). The seven basic lithologic types that display different spectral reflection characteristics were basalt, Franklin intrusive rocks, carbonate rock, evaporite, sandstone, dolostone, and Quaternary cover (Fig. 2b) The number of polygons and pixels for each training class is shown in Table 1.

The well-known maximum likelihood classification algorithm (MLC), a parametric classifier, was employed with the Robust Classification Method (RCM) to evaluate training area statistics for normal (gaussian distributions) using first-order statistics and visual analysis of histograms (both cumulative and non-cumulative). This analysis confirmed that all distributions were normal or very close to normal, thus justifying the use of the MLC algorithm. Hereafter we refer to this algorithm as RCM-MLC.

Training Area Separability

For a successful classification, the training areas must represent each rock type, be spectrally pure, and statistically separable. The separability of a pair of probability distributions can be measured by means of divergence, which is defined in terms of the likelihood ratio:

$$L_{ij}(x) = p(x|\omega_i) / p(x|\omega_j) \quad (\text{Eq. 2})$$

where $p(x|\omega_i)$ and $p(x|\omega_j)$ are the values of the i th and j th spectral class probability distributions, respectively, at position x (Richards and Jia 2006). The

separability of the training dataset was examined using the transformed divergence (TD) statistic, which is considered to be a better measure than simple divergence (Richards and Jia 2006). TD values range between 0 and 2. Values greater than 1.9 indicate that the class pairs have good separability, whereas values less than 1 indicate that the class pairs are not spectrally separable and perhaps should be combined into one class. Values between 1.5 and 1.9 indicated moderate to good separability.

Figure 3a and 3b are plots of TD for each of the thirteen lithological – surficial classes and seven generalized lithological – surficial classes (see Table 1), respectively. In general, the training areas are statistically separable for all data except for the Aster thermal imagery (thermal infrared, or TIR). The Landsat-7 and Landsat-8 data offer the best separability, closely followed by Aster, whereas Spot-5 offers lower separability. The separability results indicate that classification can proceed, although the TIR results would be marginal.

Classification

Supervised classification is based on the premise that *a priori* knowledge of the features to be classified exist (i.e. training areas); this knowledge is then used to find similar areas based on the training statistics in an n -dimensional dataset.

A number of approaches that make use of training areas for not only producing a classification but also for validating the classification are typically employed. The same training dataset can be used for producing the classification and for validating the classified map using a confusion matrix. However, using the same training dataset for both classification and validation leads to a statistical bias that will result in an inflated value of the overall accuracy (Jones and Vaughan 2010). A more robust approach is to define one training dataset for classification and another independent set for validation, or else use a specified portion of the training dataset for classification and validation. These latter approaches provide a much more robust and statistical validation of the classified map, as an independent set of training areas

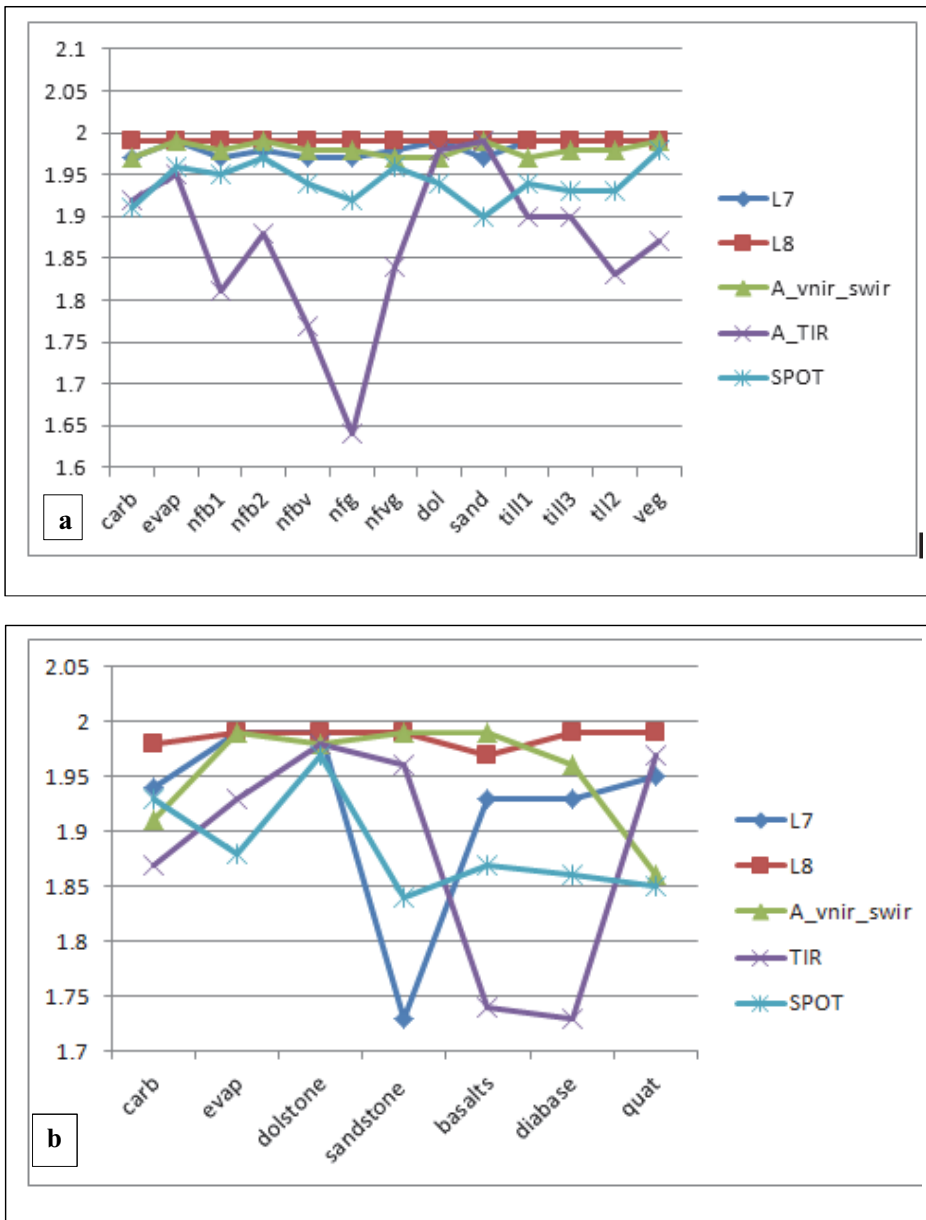


Figure 3. (a) Transformed Divergence (TD) values calculated from the training areas for the thirteen lithological – spectral classes for Landsat-7 (L7), Landsat-8 (L8), Aster (VNIR and SWIR; A_vnir_swir), Aster thermal (A_TIR), and Spot-5 (SPOT). See Table 1 (training data column) for the description of each lithological symbol along the x-axis. (b) Transformed Divergence (TD) values calculated from the training areas for the seven generalized lithologic classes for Landsat-7 (L7), Landsat-8 (L8), Aster (VNIR and SWIR; A_vnir_swir), Aster thermal (TIR) and Spot-5 (SPOT). See Table 1 (generalized lithologies and surficial materials column) for descriptions of the generalized lithotypes (along the x-axis) and how they were derived from the detailed geology (detailed lithologies and surficial materials column). For both TD plots, lithotype classes are shown along the x-axis and TD values along the y-axis.

are used to construct the confusion matrix.

Training area selection is often difficult because of the variability of spectral responses over the same terrain resulting from seasonal differ-

ences, variation in moisture content, and atmospheric effects. Furthermore, the area to be classified may be inaccessible (typical of many areas in the north) and therefore cannot always be verified by a field visit. Also, the fea-

tures to be classified (bedrock and surficial materials) may present either a limited or wide range of spectral responses, making classification difficult. In some cases there are several people involved in the selection of the training areas, also leading to variability because of different user knowledge and experience. Therefore, training area variability reflects inconsistencies in the properties inherent in the training area selection process. In this study, one of us (RR) guided the training area selection based on his knowledge of the geology and the spectral responses exhibited by different bedrock types in the remotely sensed imagery. Two ensemble classification algorithms were used to assist in quantifying the uncertainty within the training dataset.

Robust Classification Method (RCM) Classifier

The Robust Classification Method (RCM; Harris et al. 2012b) is an ensemble classifier that is based on a randomized and repeated sampling of a training dataset in concert with traditional cross-validation of the classification results. A series of predictions (classified maps) and associated uncertainty maps and statistics are produced. This method employs any supervised classification algorithm available in the ENVI™ suite of software; in this study we use the MLC. RCM is especially useful for assessing the effects of spectral and spatial variability in the classification process. Specifically, this method provides a majority classification and variability map, and ‘confusion’ statistics; these quantify the uncertainty in the classification process with respect to statistical (spectral) variability in the training dataset as well as identifying areas that show spatial variability in classification. RCM provides a more robust estimate of overall accuracy as it is based on producing a number of classifications (in this case, ten) that can be leveraged to not only provide an overall average classification accuracy but also supply an estimate of classification uncertainty (uncertainty map; see below). The RCM input parameters used herein included ten iterations, each of which sampled a random selection of 50% of the training data for classification and 50% for independent validation.

The majority classification map produced by RCM is based on the most ‘popular’ class on a pixel-to-pixel basis for the ten iterations, and the uncertainty maps are based on variability in the classification on a pixel-to-pixel basis. Average rule images that show the strength of membership of each pixel to a specific class, similar to a class probability image, are also generated. Harris et al. (2102b) provide a more detailed description of the RCM approach. RCM using the maximum likelihood classification algorithm (RCM–MLC) was applied to the twelve combinations of remotely sensed data shown in Table 4. Two sets of classifications were produced, based on the above data combinations: one for the thirteen spectral unit classes and one for the seven generalized (bedrock and surficial) classes (Table 1) to facilitate a comparison with the generalized geology map (Fig. 2b).

Random Forests (RF) Classifier

Random Forests (RF) is an ensemble, and multiple decision-tree classifier that offers a number of advantages for classification:

- data can be binary, categorical or continuous;
- the classifier performs internal cross-validation through ‘bootstrapping’, which provides a robust estimate of classification accuracy using *out-of-bag* estimates;
- it is a non-parametric classifier and is relatively insensitive to outliers in the training data;
- it requires little user input (m , the number of decision trees, and n , the number of variables for each decision tree);
- it produces a classification map, but more importantly, probability maps (strength of membership in each lithological class); and
- it ranks the input variables with respect to their importance in the predictions.

Random Forests was originally developed by L. Breiman and A. Cutler at the University of California, Berkeley (Breiman et al. 1984; Brieman 2001). Training data are required for this approach, similar to other supervised classifiers. For each tree (the number of decision trees, m , is determined by the operator), a random

Table 4. Data combinations used for classification.

| | |
|-------------|---|
| Landsat-7 | 6 spectral bands – 3 visible, 1 NIR and 2 SWIR |
| Landsat -8 | 7 spectral bands – 4 visible (two blue bands), 1 NIR and 2 SWIR |
| Aster | 9 spectral bands - 2 visible, 1 NIR and 6 SWIR |
| Aster | 5 TIR bands |
| Spot-5 | 4 bands – 2 visible, 1 NIR and 1 SWIR |
| Landsat -7 | 6 spectral bands + 4 ratios (see Table 3) |
| Landsat- 8 | 7 spectral bands + 4 ratios (see Table 3) |
| Aster | 9 spectral bands + 12 ratios (see Table 3) |
| Aster | 7 – 6 spectral bands + total field magnetic data |
| Landsat- 8 | 7 spectral bands + total field magnetic data |
| Landsat – 7 | 6 spectral bands and 4 ratios + total field magnetic data (for generalized classification only) |
| Landsat - 8 | 7 spectral bands and 4 ratios + total field magnetic data (for generalized classification only) |

selection of the input variables (i.e. remotely sensed image bands, n) is made. The number of variables selected for each tree is a fraction of the total number of variables; the square root of the number of variables is often used. Each tree employs a ‘bagging’ process (i.e. ‘bootstrap’ sample; Brieman 1996) whereby approximately two-thirds of the training areas (pixels) are used to create a prediction (referred to as *in-bag*) and one-third to validate the accuracy of the prediction (referred to as *out-of-bag*, or *oob*). This random sampling with replacement of the training dataset is undertaken for every tree. *In-bag* data are used to create multiple decision trees that are applied to produce independent classifications. At each node of the individual decision tree, the best split is chosen from a random sample of variables. Each tree is grown to the maximum extent with no pruning. We used the *Gini* index to determine the impurity at each node:

$$\text{Gini Index} = 1 - \sum c(p^2(c|t)) \quad (\text{Eq. 3})$$

where c = number of classes (e.g. lithotypes), t = node of a tree, p = relative frequency of c (a given lithotype class)

The stop criteria for splitting each node is based on the minimum of samples in a node (we used 1) and the

minimum impurity in a node (we used 0) allowing full growth of the decision tree (no pruning). Thus, an ensemble of trees (predictions) is created and a voting procedure is employed to assign the majority class to each pixel in the final prediction map. According to Brieman (2001), Gislason et al. (2006) and Menze et al. (2009), RF is not sensitive to noise or over-fitting and there is no need for cross-validation as it is estimated internally. However, as with any supervised classification method, an independent check of the training dataset of each lithotype is still required to calculate an unbiased and more robust estimate of classification accuracy. Additionally, the probability of membership in each class is also generated, which can be used to assess the uncertainty of the RF classification. Figure 4 summarizes the RF classification process.

Another very useful aspect of RF is that it calculates the importance (predictive power) of each variable in the classification process. This is accomplished by:

- For each tree the *oob* samples are permuted in the respective variable and then put down the tree and the number of correct classifications are calculated (nP);
- The *in-bag* training samples (original) are put down the tree and the

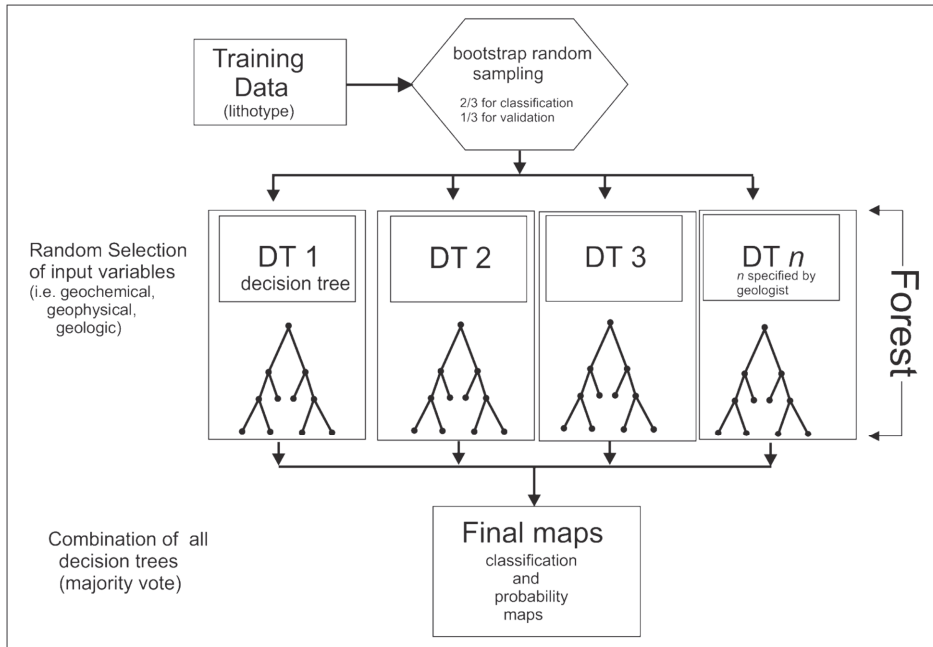


Figure 4. Summary of the Random Forest (RF) classification algorithm.

number of correct classifications are calculated (nC); $nC-nP$ calculated;

- Calculate $Nc-Np$.

The average of the differences of the accuracies for all trees is the raw importance of each variable. However, to provide a more robust estimate the raw variable importance is divided by the respective standard deviation creating a normalized variable importance value. A high normalized value has a high importance for the entire RF and *vice versa* for a low number.

A RF classification map using all the available raw and ratio data from the Landsat-7 and Landsat-8, Aster (VNIR: visible near infrared; SWIR: shortwave infrared; and TIR) and magnetic data, totalling 43 input bands, was produced along with a probability-of-membership map (same as a rule map for RCM-MLC), which was used to reduce the uncertainty by only accepting pixels for each class that showed a strong membership (>95%). In addition, the 43 input bands were ranked for their predictive power using the method as discussed above. The RF parameters used include n (number of variables to create each tree), which was set to the square root of the total number of variables, and m (number of decision trees), which was set to 100 through experimentation. As pointed out by Rodriguez-Galiano et al. (2014),

the number of trees is proportional to the classification accuracy until the *oob* error converges. Once the error converges at a given number of trees, the number of random variables selected for each tree (n) minimally alters the classification accuracy. In this study, the *oob* error stabilized at 100 decision trees (m) in all of our experiments.

Ensemble Classifications

Ensemble methods (Benediktsson and Swain 1992; Doan and Foody 2007; Benediktsson et al. 2007; Banfield et al. 2007) leverage multiple models (i.e. classifications) to obtain a better prediction than could be derived from individual predictions. The flexibility of ensembles can, in theory, enable them to over-fit the training data more than a single prediction would, but in practice, some ensemble techniques, especially bagging (Breiman 1996; Freund and Schpire 1999) tend to reduce problems related to over-fitting of the training data. In addition, they provide a more robust estimate of the overall classification accuracy. Both classification algorithms used in this study, RCM-MLC and RF, are ensemble classifiers as they produce multiple predictions (classifications), which are combined using a voting procedure that selects the most popular class on a pixel basis.

Multiple Classification Systems (MCS)

Multiple classification systems (MCS; Polikar 2006; Benediktsson et al. 2007; Waske and Braun 2009) are a fairly recent concept; they do not refer to a specific classification algorithm but describe a more general classification strategy. MCS can involve combining classifications based on a number of different algorithms, variants of the same algorithm, or classifications derived from different datasets. In this case, the latter approach was utilized by combining classifications derived from Landsat, Spot-5, Aster and airborne magnetic data. Specifically, we use a *stacking* approach whereby the data combinations that produced the highest classification accuracies for the generalized lithological classes (seven in total) were combined in an ensemble fashion by (a) taking the majority class (most ‘popular’) for each pixel (*MCS majority map*), and (b) only classifying a pixel when the class was in agreement with all input classifications comprising the stack (*MCS agreement map*).

As part of this MCS process, and because multiple classifications are being combined, uncertainty measures such as the range and/or standard deviation of the classification on a stacked pixel basis may be calculated. In this study, a standard deviation calculation was used on the pixel stack (derived from the 100 decision trees for RF and 10 iterations for RCM-MLC) to estimate uncertainty in the MCS classification process. The standard deviation map was then thresholded and used to mask-out uncertain areas on the MCS classification maps.

A variation of the MCS theme was also added by selecting the data combination that provided the highest classification on a class-to-class basis and producing a classification map from these selections. Thus, this *MCS best bands majority classification map* contains the class from the data combination that provided the highest classification accuracy. To further decrease the uncertainty in this classification image, the average *rule or probability images* (strength-of-class membership, discussed above) generated by RCM-MLC were used to create the classification map. Each *rule image* for each data combination that provided the highest

classification accuracy was thresholded at the 95% level, thereby presenting, for each class, the pixels having the strongest membership. Again, this produced a map where only the most certain pixels were classified.

CLASSIFICATION AND RESULTS

RCM–MLC Classifications: Thirteen Classes

Table 5 presents the results of the RCM–MLC classification for all data combinations for all thirteen lithological – surficial classes (Table 1), and Figure 5 is a box and whisker plot of the average accuracies over the ten iterations of RCM–MLC for each data combination. Figure 6 is a plot of the average producer's and user's accuracy. Figure 7a shows the majority classification map for the thirteen classes

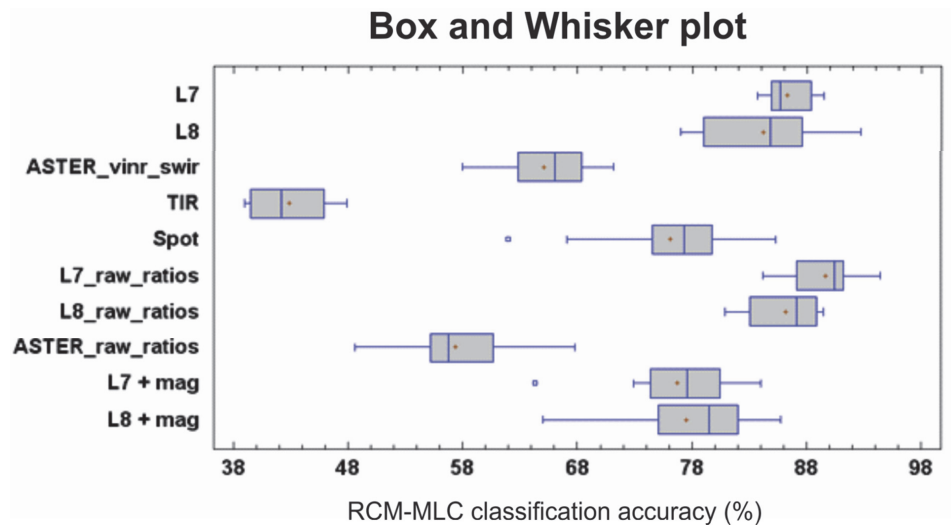


Figure 5. Box and whisker plot showing average (shown as a cross symbol) and median (shown as a vertical line) classification accuracies as well as the range in accuracies for ten iterations of Robust Classification Method–maximum likelihood classification (RCM–MLC) (using various data combinations) for thirteen lithological – spectral classes.

Table 5. Results of Robust Classification Method–maximum likelihood classification (RCM–MLC) classification showing average overall, producer's and user's accuracies and average class accuracy for each data combination for each of the thirteen lithological – spectral classes (Table 1). Bold numbers are the highest classification accuracies.[^]

| Class | L7_r | L8_r | L7_rr | L8_rr | ASTER VNIR SWIR_r | ASTER -TIR | ASTER_rr | L7_rr_mag | L8_rr_mag | SPOT |
|-----------------------------|-------------|-------------|-------------|-------------|-------------------------|---------------|----------|-----------|-----------|------|
| Carbonates | 85.5 | 97.9 | 86.2 | 94.9 | 85.4 | 65 | 59.3 | 96 | 98 | 79 |
| Evaporites | 87.3 | 98.2 | 89.7 | 87.3 | 89.4 | 60.9 | 51.2 | 90.8 | 84.7 | 91.5 |
| Basalts Nfb1 | 79.3 | 55.9 | 87.9 | 67.1 | 11.9 | 12.2 | 25.2 | 54.3 | 40.2 | 76.7 |
| Basalts Nfb2 | 92.6 | 72 | 86.7 | 57.1 | 41 | 23.7 | 56.8 | 91.1 | 77.9 | 78.8 |
| Basalts Nfbv | 74.6 | 88.2 | 75.8 | 83.8 | 81.9 | 24.6 | 58.9 | 69.1 | 66.3 | 74.1 |
| Franklin Diabase Nfg | 91.2 | 92.3 | 92.6 | 95.2 | 81.9 | 36 | 83.7 | 91.2 | 87.8 | 87.6 |
| Franklin Diabase Nfgv | 73.7 | 92.2 | 89 | 94.2 | 21.2 | 7 | 41 | 47.7 | 72.8 | 82.5 |
| Dolostones | 83.2 | 95.3 | 80.1 | 96.4 | 56.7 | 15.1 | 65.2 | 65.2 | 74.6 | 74.8 |
| Sandstone | 61.7 | 78 | 84.6 | 74.6 | 62.3 | 93.8 | 32.8 | 50.7 | 58.7 | 47.1 |
| Till1 | 90.5 | 79.1 | 93.5 | 75.3 | 81.3 | 70 | 79.5 | 76.5 | 64.3 | 62.1 |
| Till2 | 92.6 | 78.9 | 97.6 | 91.8 | 60.9 | 20.2 | 53.2 | 73 | 75.9 | 57.8 |
| Till3 | 89.4 | 88 | 92.8 | 96.3 | 79.3 | 32.7 | 75 | 95.4 | 86.7 | 71.8 |
| Vegetation | 97.2 | 70.6 | 95.8 | 91.1 | 45.3 | 16.4 | 13.4 | 82 | 80.6 | 90.5 |
| Overall accuracy | 86.3 | 84.2 | 89.7 | 86.2 | 65.1 | 42.9 | 57.3 | 77.1 | 77.5 | 76.1 |
| Average Producer's accuracy | 85.5 | 83.1 | 88.4 | 84.6 | 61.3 | 40 | 54.5 | 81.3 | 74.9 | 74.3 |
| Average Users Accuracy | 87.3 | 86.6 | 90.1 | 87.4 | 64.6 | 42.5 | 57.6 | 86.3 | 81.2 | 75.4 |

[^] Key to abbreviations: L7_r = Landsat-7 (raw); L8_r = Landsat-8 (raw); L7_rr = Landsat-7 (raw and ratios); L8_rr = Landsat-8 (raw and ratios); ASTER VNIR_SWIR_r = Aster very near infrared and shortwave infrared (raw); ASTER–TIR= Aster thermal bands; ASTER_rr = Aster VNIR and SWIR (raw and ratios); L7_rr_mag = Landsat-7 (raw, ratios and magnetics); L8_rr_mag = Landsat-8 (raw, ratios and magnetics); SPOT = Spot-5 (raw).

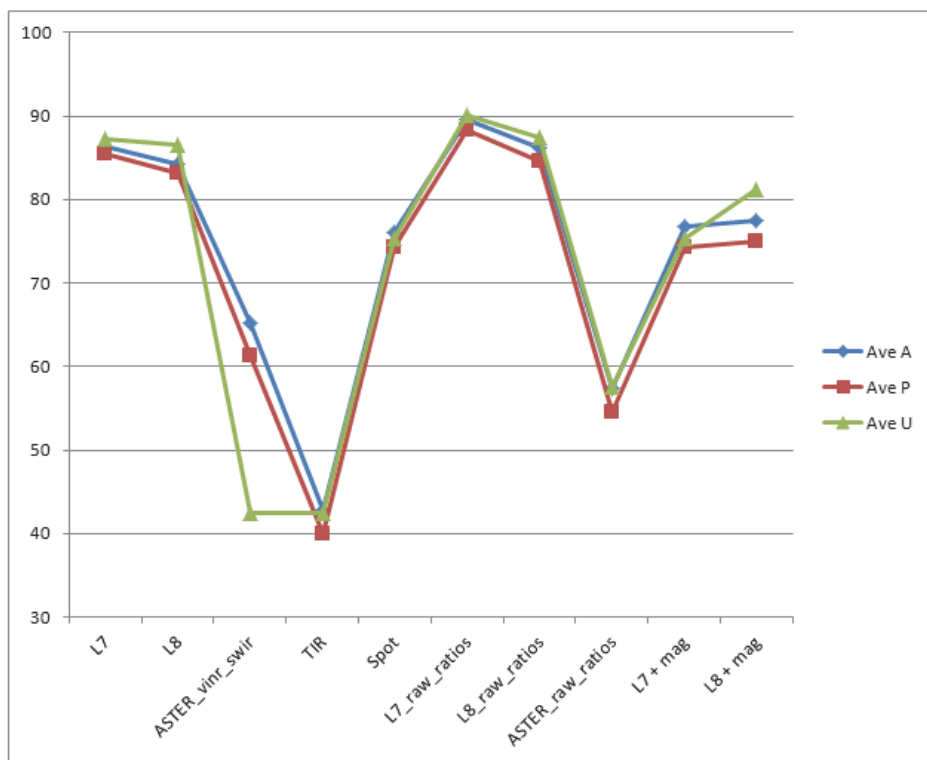


Figure 6. Plot of average overall accuracy (Ave A), average Producer’s accuracy (Ave P) and average User’s accuracy (Ave U) for ten iterations of the Robust Classification Method–maximum likelihood classification (RCM–MLC; using various data combinations) for thirteen lithological – spectral classes. The data combination is shown on the x-axis and classification accuracy (0–100%) along y-axis.

derived from the data combination that produced the overall RCM–MLC highest classification accuracy (Landsat-7 raw and ratios data – 89.7%) and Figure 7b shows the associated RCM–MLC uncertainty map.

With respect to RCM–MLC, the highest classification accuracies were obtained from Landsat-7 (raw and ratios), Landsat-7 (raw), Landsat-8 (raw and ratios), and Landsat-8 (raw), all of which have an overall accuracy greater than 80%. The other combinations, including Aster (especially the TIR) and Spot-5, have lower classification accuracies. The results from Aster are generally lower and no doubt relate to the acquisition parameters (Table 2), primarily a lower sun angle, which would result in a lower signal-to-noise ratio.

With reference to Table 5, the carbonate rocks are classified most accurately on the Landsat-8 (raw and magnetics) data, as these rocks have very low magnetic susceptibility (compared to the basalts and Franklin diabase) and very high reflectance in the visible bands and band 6 (SWIR), but

low reflectance in band 7 (SWIR). Note that the accuracies for carbonate are consistently high for all data combinations with the exception of Aster. Evaporites are also classified with high accuracy on all data combinations, with the exception of Aster (both VNIR–SWIR and TIR); Landsat-8 (raw data) provided the highest classification accuracy (98.2%).

The Natkusiak basalts, which comprise three distinct spectral sub-units (see Table 1), are best classified on combinations of the Landsat-7 and Landsat-8 data. Landsat-7 (raw and ratios) provide the highest classification accuracy for Nfb1 (87.9%), Landsat-7 (raw) for Nfb2 (92.6%) and Landsat-8 (raw) for Nfbv (88.2%). It is interesting to note that addition of the magnetic data does not provide higher classification accuracies compared to the Landsat-7 and Landsat-8 optical data used alone. The Franklin intrusive rocks (diabase), which comprise two spectral sub-units (Table 1), are best classified on the Landsat-8 (raw and ratio) data combination (Nfg – 94.2%; Nfgv – 94.2%). Inclusion of the Landsat-8

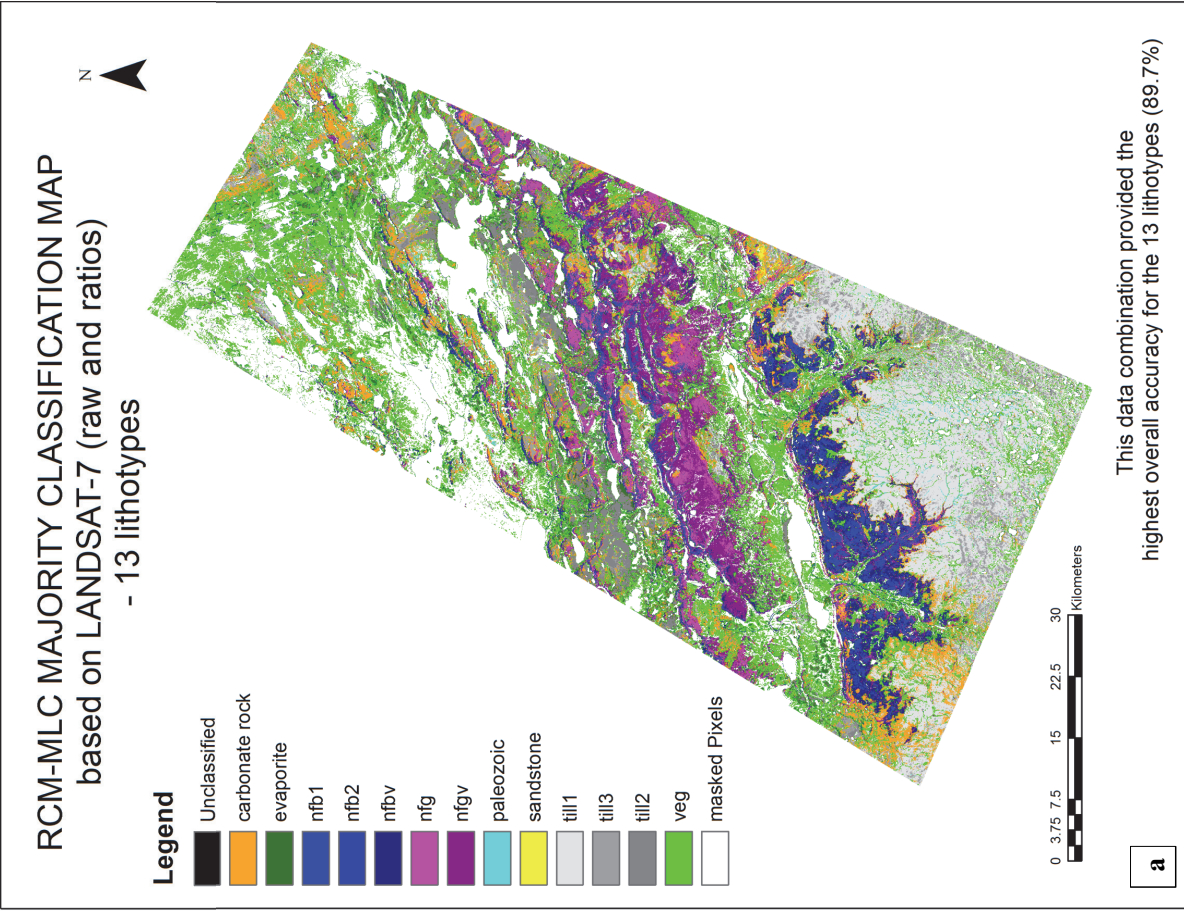
ratios improves the overall classification accuracy from 3 to 5% for the Nfg and Nfgv sub-units, respectively. Dolostones are again best classified on the Landsat-8 data, especially when the ratios are included (96.4%).

In general, the Aster TIR data offer little separation between the different rock types; however, the TIR data provide appreciably higher classification accuracy for sandstone (93.8%) than the other data combinations. The data combination that offers the second highest classification for sandstone (84.6%) is the Landsat-7 (raw and ratios). Visual examination of the thermal data indicates that the training areas used for classification of sandstone are thermally much cooler than other rock types, especially at the 8.125, 8.475 and 8.925 μm wavelengths. This is also confirmed by the separability values (TD) (Fig. 3a), which reveal that sandstone has the highest TD value for all thirteen classes.

The Landsat-7 and Landsat-8 (raw and ratios) data provide the highest classification accuracies for the three spectrally distinct till classes. Again, inclusion of the ratio data increases the accuracies by an average of 4%. The Aster and Spot-5 data provide lower classification accuracies when compared to the Landsat data. It is interesting to note that inclusion of magnetics with the Landsat-7 (raw and ratio) data increases the classification accuracy from 92.8 to 95.4%, very close to the best accuracy provide by Landsat-8 (raw and ratio; 96.3%). Visual analysis of the magnetic data indicates that till unit 3 has a lower magnetic signature than the others, which may reflect lower magnetic susceptibility of the underlying bedrock (not known because of glacial cover), and/or a lower magnetite content in the overburden compared to other till units.

RCM–MLC Generalized Lithological Classifications: Seven Classes

Figure 8 is a box and whisker plot of the average RCM–MLC accuracies for nine data combinations (Table 4) based on the seven generalized lithological – surficial classes (Table 1). Table 6 provides a summary of overall accuracy, producer’s and user’s accuracy, as well



This data combination provided the
highest overall accuracy for the 13 lithotypes (89.7%)

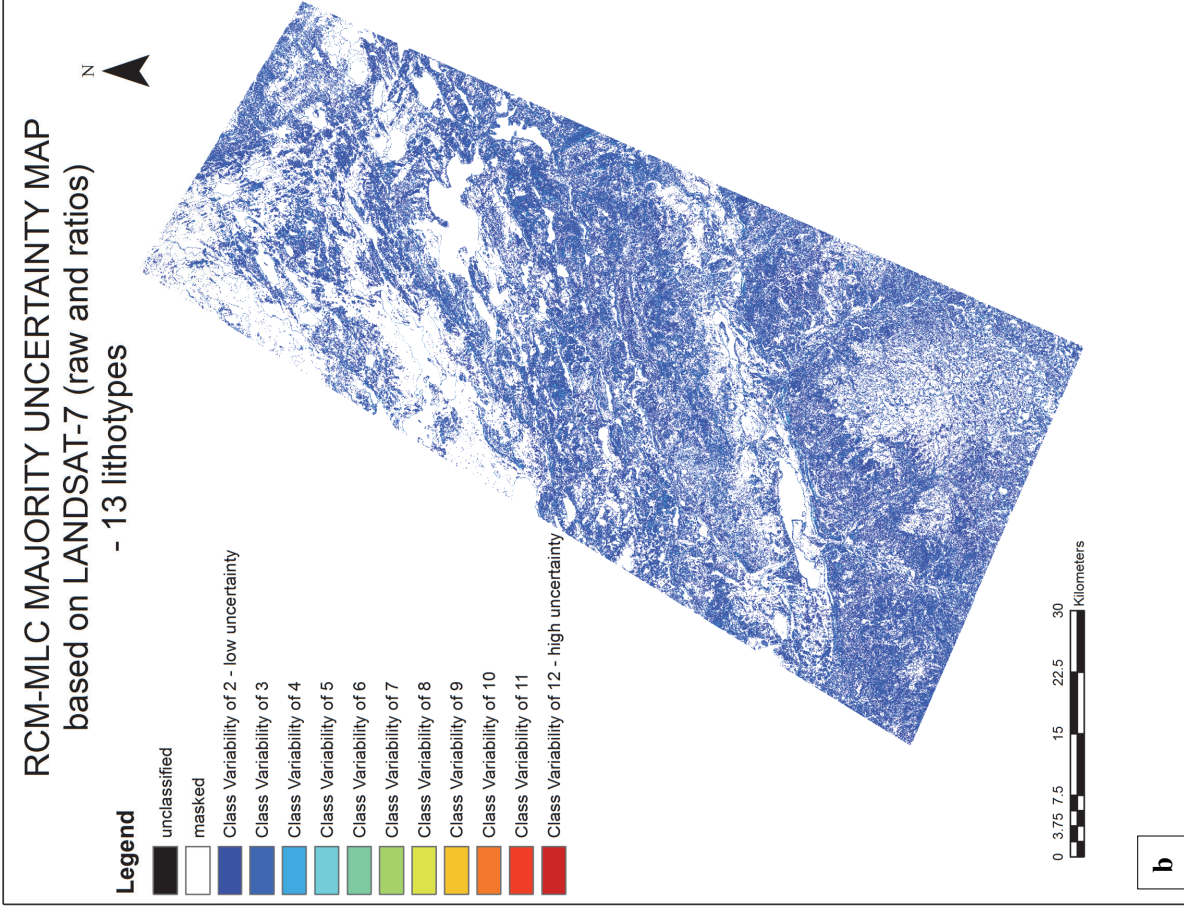


Figure 7. (a) Robust Classification Method–maximum likelihood classification (RCM–MLC) Majority Classification Map of thirteen classes, produced from the Landsat-7 (raw and ratios) data combination that produced the highest overall accuracy (89.7%) of all data combinations. See Table 1 for description of lithologic units. (b) RCM–MLC uncertainty map. Note that few pixels are uncertain; hence warmer colours are difficult to see on the uncertainty map.

Box and Whisker plot

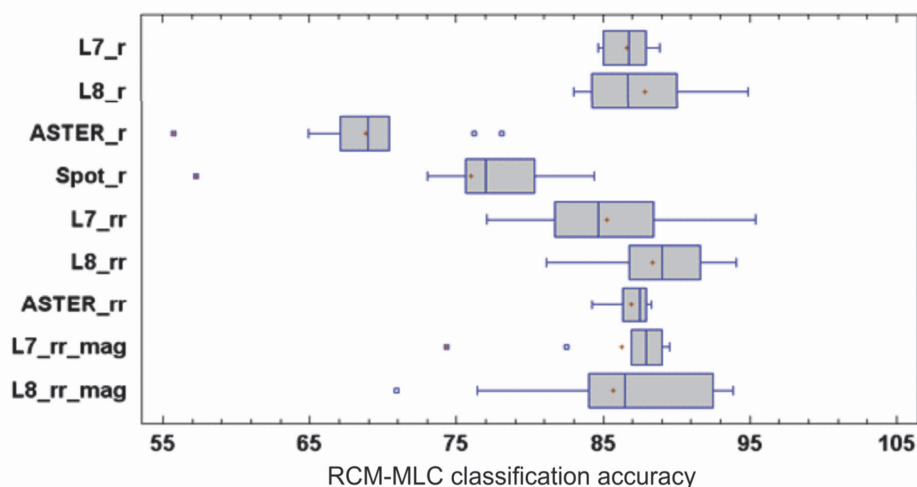


Figure 8. Box and whisker plot showing average (shown as a cross symbol) and median (shown as a vertical line) classification accuracies, as well as the range in accuracies for the ten iterations of the Robust Classification Method–maximum likelihood classification (RCM–MLC) for the seven generalized lithological – surficial classes (see Table 1).

as classification accuracy by class for all the data combinations. Note that TIR was included with the Aster (raw and ratio) data combination mainly because of its ability to discriminate sandstone. Also note that for the generalized clas-

sifications two more data combinations were added – Landsat-7 and Landsat-8 (raw, ratios, and magnetic data) (Table 4). These were included because of the additional value the ratio and magnetic data provided for classifying the thir-

teen lithological – spectral classes, discussed above. The data combinations comprising the Landsat-7 and -8 (raw), Landsat-7 and -8 (raw and ratios), Landsat-7 and -8 (raw, ratios and magnetics) and the Aster (raw and ratios) data offer classification accuracies in excess of 85%; the Landsat-8 (raw and ratios) offers the highest overall, producer’s, and user’s accuracies of 88.4, 87.7 and 90.1%, respectively (Table 6). Figure 9a shows the majority classification map for the seven generalized lithological – surficial classes derived from the data combination that produced the highest overall classification accuracy (Landsat-8 raw and ratios – 88.4%), and Figure 9b shows the associated uncertainty map. The Aster (raw) and Spot-5 data combinations resulted in significantly lower accuracies. In general, the classification accuracies are higher for the generalized lithologies (seven classes) versus the more detailed lithologies and thirteen spectral sub-units for all data combinations. This is expected because of the reduction and generalization of the detailed lithological classes.

Table 6. Results of Robust Classification Method–maximum likelihood classification (RCM–MLC) classification showing average overall, producer’s and user’s accuracy and average class accuracy for each data combination for each of the seven generalized lithological – surficial classes. Bold numbers are the highest classification accuracies.^

| Class | L7_r | L8_r | L7_rr | L8_rr | ASTER VNIR SWIR_r | ASTER_rr_TIR | L7_r_mag (L7_rr_mag) | L8_r_mag (L8_rr_mag) | SPOT-5 |
|-----------------------------|-------------|-------------|-------|-------------|-------------------|--------------|----------------------|----------------------|--------|
| Carbonates | 87 | 96.2 | 89 | 93.3 | 81.5 | 69.6 | 80.8 (92.2) | 91.8 (95.3) | 78.9 |
| Evaporites | 82.4 | 82.4 | 89.2 | 84.4 | 73.7 | 89.6 | 95.5 (98.6) | 85.4 (95.3) | 93.8 |
| Basalts | 90.1 | 87.8 | 87.5 | 81.4 | 63.4 | 90.9 | 90.7 (86.4) | 80.2 (76.1) | 71.2 |
| Franklin diabase | 85.1 | 87.4 | 79.9 | 86.4 | 68.5 | 88.9 | 80.4 (82.8) | 82.8 (90.5) | 72.9 |
| Dolostones | 91 | 89.7 | 87.9 | 96.8 | 68.5 | 73.2 | 72 (53.8) | 88.9 (91.2) | 83.4 |
| Sandstones | 62.9 | 70.6 | 62.7 | 73.1 | 30.1 | 86.7 | 42.2 (77.8) | 72.7 (41.9) | 66.4 |
| Quaternary – till cover | 92.7 | 90.4 | 92.6 | 94.9 | 77.8 | 91.8 | 93.9 (91.5) | 93.2 (92.4) | 79.5 |
| Overall accuracy | 86.6 | 87.9 | 85.3 | 88.4 | 68.8 | 86.9 | 83.2 (86.3) | 86.1 (85.7) | 76 |
| Average Producer’s accuracy | 83.8 | 85.8 | 84.5 | 87.7 | 66.7 | 84.3 | 79.8 (83.1) | 85 (83.7) | 78.1 |
| Average Users Accuracy | 86 | 89.6 | 87.3 | 90.1 | 65.8 | 86.2 | 81.7 (87.7) | 88.1 (87.1) | 73.3 |

^ Key to abbreviations: L7_r = Landsat-7 (raw); L8_r = Landsat-8 (raw); L7_rr = Landsat-7 (raw and ratios); L8_rr = Landsat-8 (raw and ratios); ASTER VNIR SWIR_r = Aster very near infrared and shortwave infrared (raw); ASTER_rr_TIR = Aster, all bands, ratios and thermal bands; L7_r_mag = Landsat-7 (raw, magnetics); L7_rr_mag = Landsat-7 (raw, ratios and magnetics); L8_r_mag = Landsat-8 (raw, magnetics); L8_rr_mag = Landsat-8 (raw, ratios and magnetics); SPOT-5 = Spot-5 (raw).

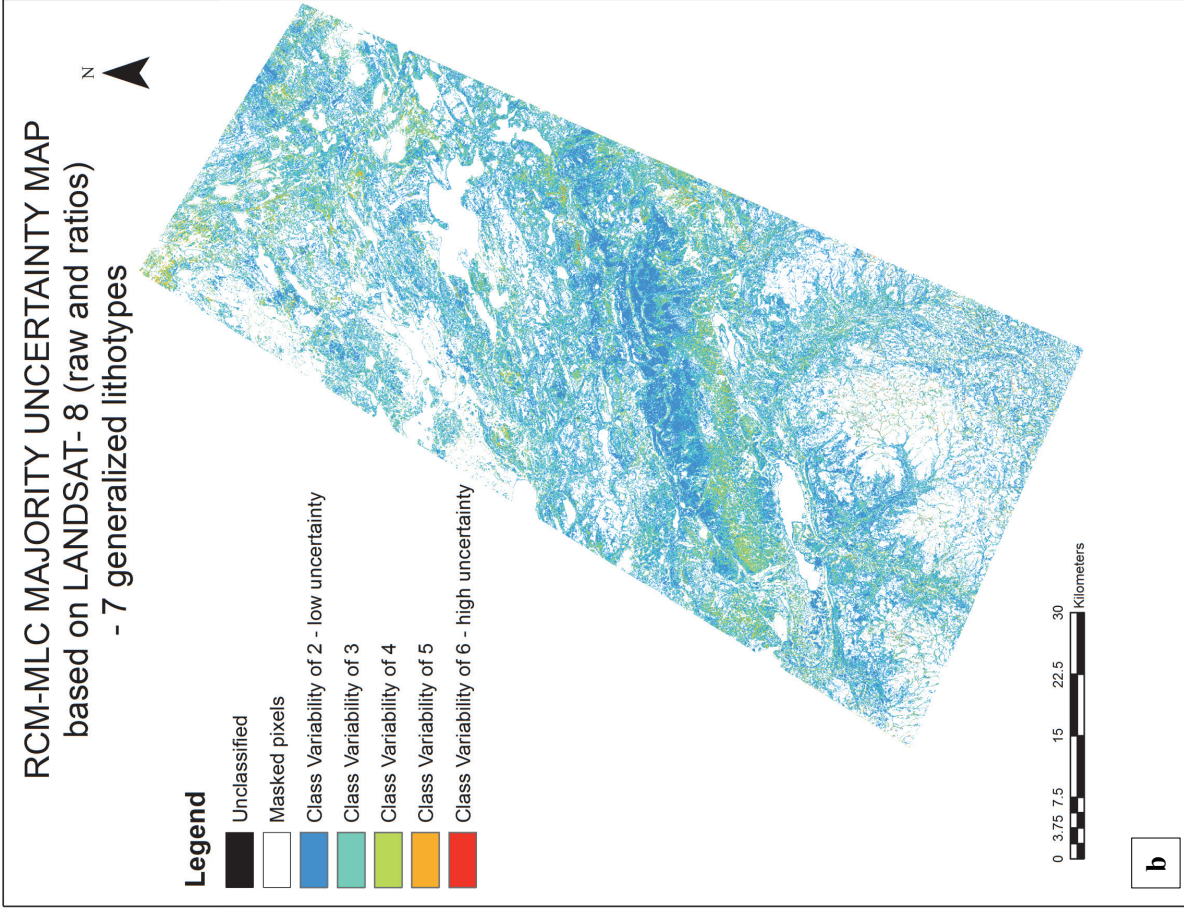
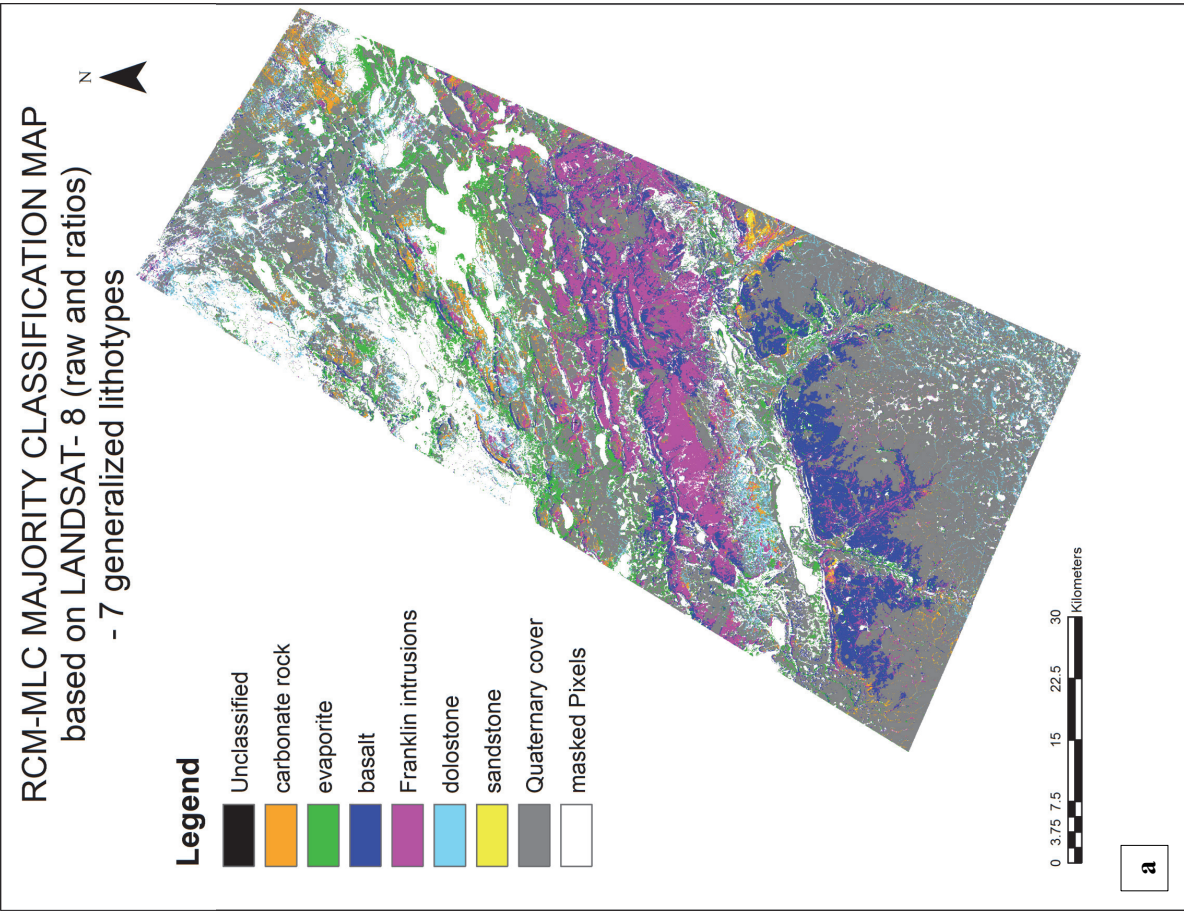


Figure 9. (a) Robust Classification Method–maximum likelihood classification (RCM–MLC) Majority Classification Map of seven generalized lithological – surficial classes, produced from the Landsat-8 (raw and ratios) data combination that produced the highest overall accuracy (88.4%) of all data combinations. (b) RCM–MLC uncertainty map.

The carbonate rocks, once again, are most accurately classified on the Landsat-8 (raw; 96.2%) and the Landsat-8 (raw, ratios, magnetics; 95.3%) data combinations (Table 6). Addition of the ratios and magnetic data in this case does not improve the overall accuracy. Again, recall that carbonates are easy to discriminate using optical data. The evaporites are classified most accurately (98.6%) on the Landsat-7 (raw, ratios, magnetics) data combination, although they have a very low magnetic susceptibility. Addition of the ratio data again helps to increase the overall accuracy by approximately 6% and the magnetic data by approximately 5%. The basalts are best classified using, suprisingly, the Aster (raw and ratios), Landsat-7 (raw, magnetics) and Landsat-7 (raw), all offering greater than 90% accuracy. In this case, the ratio data do not improve classification accuracies. The magnetic data are useful because the basalts have an enhanced magnetic response (higher magnetic susceptibility). The Landsat-8 (raw, ratio, magnetics) data combination results in the highest classification accuracy for the Franklin intrusive rocks. Addition of the ratios and the magnetic data, because of their high magnetic response, offers an increase in accuracy of approximately 4%. The dolostones are best classified by the Landsat-8 (raw and ratios) data combination, which is 5% higher than any other data combination. Inclusion of the ratios improved the accuracy by approximately 7% compared to the raw data only. The Aster data provided the highest classification accuracy for sandstones (VNIR and SWIR; raw and ratios), for reasons relating to the thermal data, discussed above. The accuracy is on the order of 15% higher than for any other data combination. Again, the Landsat-8 (raw and ratios) data provided the best classification of Quaternary deposits (94.9%).

Random Forest (RF) Classification

Figure 10a shows the RF classification map and Figure 10b is a RF map in which uncertain pixels have been removed, as discussed in a previous section. Overall, the RF map produced the highest classification accuracy of 99%, which exceeds the highest overall RCM-MLC classification accuracy, that

of the Landsat-8 (raw and ratios) data combination (88.4%), by approximately 10%. However, one has to keep in mind that all 43 bands (raw and ratios) from the Landsat-7, Landsat-8, and Aster (VNIR, SWIR, TIR) were used for classification. The RCM-MLC classifications used various data combinations, discussed above, but not the total number of bands.

The other purpose of using the RF classifier was to rank the predictive power of each input band, the results which are shown on Figure 11 (*Gini* index) and Table 7. The groups of variables shown in Table 7 correspond to the breakpoints as indicated in the *Gini* index plot (Fig. 11). With respect to the ranking of the bands that provide the most predictive power, the ferrous Landsat-7 ratio and Landsat-8 band 5 (NIR) (Group 1 – Table 7 and Fig. 11) provide the most significant contributions. Group 2, also significant contributors, contain two Landsat-8 ratios, ferrous iron and biomass. Group 3 primarily comprises SWIR and NIR bands derived from Landsat-7 and Landsat-8. Of interest is the Aster NIR band, also included in this group, which provides moderate predictive power. Also of note are the total field magnetic data found in Group 6. The ferrous ratios and magnetic data are useful predictors, especially for the basalt and Franklin intrusive rocks, which are characterized by ferrous iron minerals and magnetite that result in high magnetic susceptibilities. However, the classes characterized by low magnetic susceptibility can also be separated using the magnetic data, as discussed above.

Multiple Classification System – Seven Classes

Figure 12 shows the *MCS majority classification map*, discussed above, in which each pixel was classified with the majority class, from the ensemble stack of data combinations that provided an overall classification accuracy in excess of 80% (Table 6). These included Landsat-7 (raw), Landsat-8 (raw), Landsat-7 (raw and ratios), Landsat-8 (raw and ratios), Landsat-7 (raw, ratios and magnetics), Landsat-8 (raw, ratios and magnetics) and Aster (raw, ratios and TIR).

Figure 13a is an ensemble map that shows the class agreement (*MCS classification agreement map*) for the best seven classifications of the generalized lithotypes (>80% overall accuracy; Table 6), whereas Figure 13b shows the uncertainty associated with this map. Recall that the uncertainty map is based on the standard deviation of the majority vote for each pixel. Figure 13c is the *MCS agreement map* with the uncertainty, composed of the pixels in class 5 and 6 (most uncertain = less agreement; Fig. 13b) removed. Note that not all the pixels in Figure 13c have been classified, because uncertainties in the classification process associated with snow, ice, high biomass, and pixels that do not show agreement, as well as low reflection from back slopes, have all been removed.

The data combination that provided the highest classification accuracy for each of the seven generalized lithological classes (Table 8) was used to produce a *MCS best bands majority classification map* (Fig. 14). In addition, the *average rule images* were used for each class to include only pixels that have a membership greater than 95% probability, as discussed above. For sandstone, which had the lowest classification accuracy, the best classification was achieved by the Aster (raw, ratios and TIR) data combination, whereas evaporites had the highest classification accuracy produced from the Landsat-7 (raw, ratios and magnetics) data combination.

Figure 15 is a *MCS map* that combines the three *MCS maps* discussed above – the *MCS majority classification, agreement and best bands maps*. This was produced by first taking pixels that were in agreement on all three maps and then pixels that were in agreement on a combination of two of the three maps. As shown in the legend, only two classes, basalt and carbonate rocks, were in agreement on all three *MCS maps*; the others were in agreement on only two of the *MCS maps*. This indicates that the basalts and carbonate rocks were easiest to classify as they have unique spectral and magnetic signatures.

Comparison to Generalized Geology Map

A comparison of the generalized

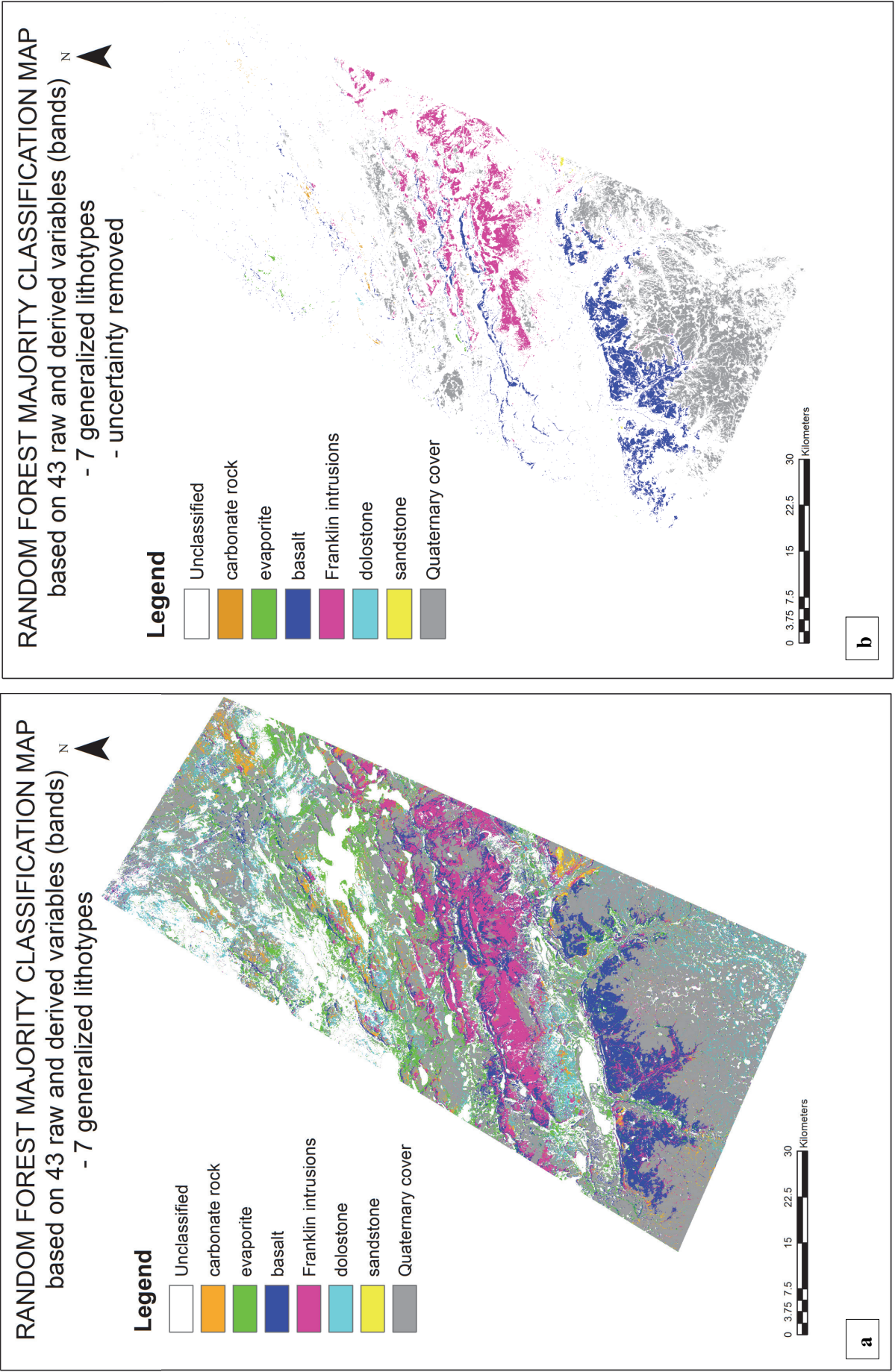


Figure 10. (a) Random Forest (RF) classification map of seven generalized lithological – surficial classes using all 43 input bands; an overall accuracy of 99% was achieved. (b) Associated RF classification map with uncertainty removed.

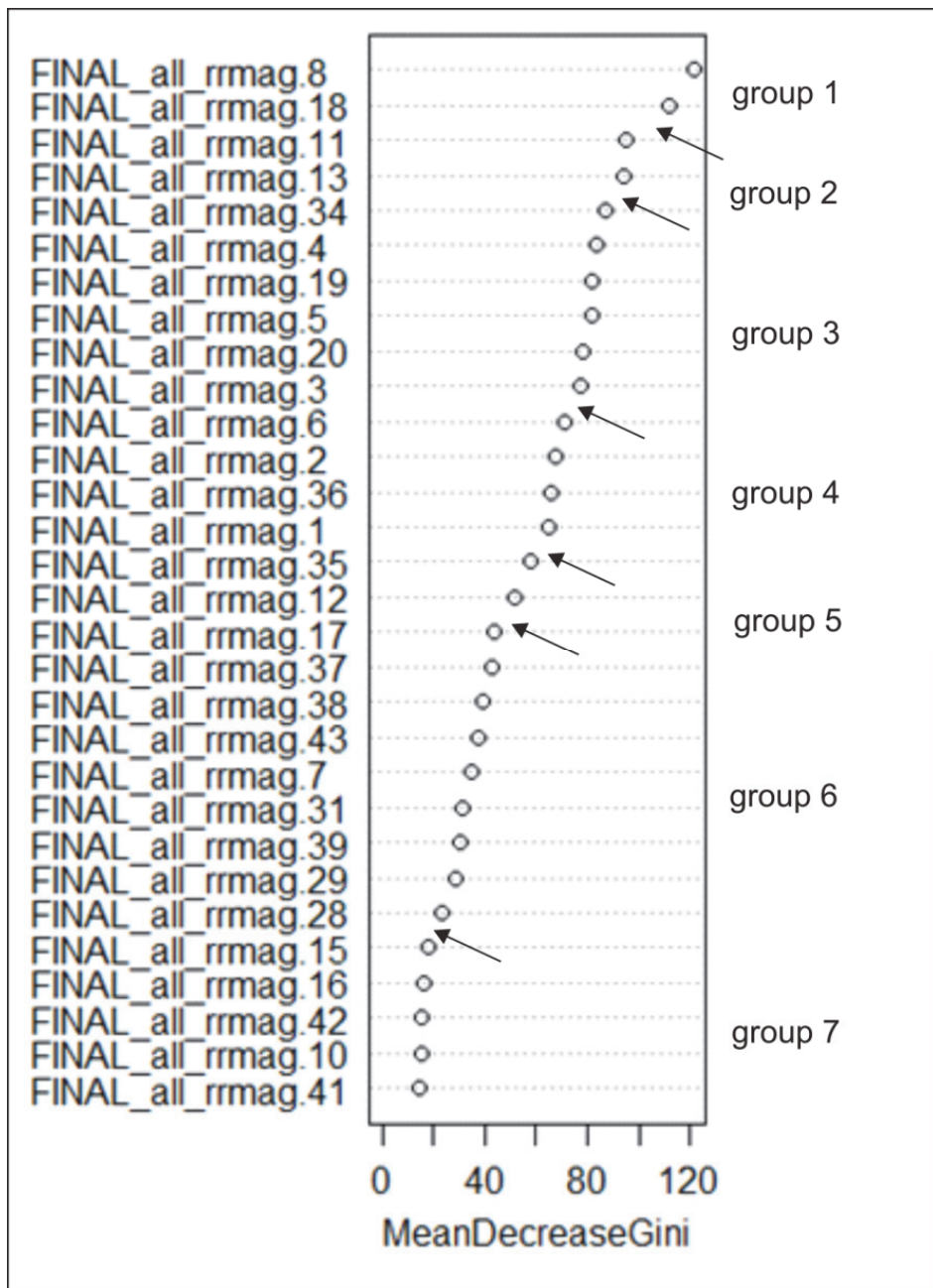


Figure 11. Gini index plot showing variable (band) importance. Arrows indicate breakpoints that divide the bands into groups. See Table 7 for description of bands composing each group.

lithology (Fig. 2b) with the generalized classification maps (seven generalized lithological classes) generated by both RCM-MLC and RF was undertaken, as discussed in a previous section. The generalization was necessary, as spectral heterogeneity recorded on the more detailed classification maps (thirteen classes) occurred within each of the generalized lithological units shown on Figure 2b. Table 9 summarizes the classification agreement with the generalized lithology. The data combina-

tion that provided the best agreement when compared to the generalized geology map was achieved by Landsat-8 (raw, ratios and magnetics, 46.1%) followed by the Landsat-7 (raw, ratios and magnetics, 44.9%). The RF classification (Fig. 10), using all 43 input bands, resulted in an agreement of 42.4% when compared to the generalized geology map even though it provided the highest classification accuracy of 99% using cross-validation (*oob*) of the training areas. It is clear that

addition of the ratios and magnetics data has improved the overall accuracy when compared to the generalized lithology map.

The *MCS best bands* classification (RCM-MLC; Fig. 14) offers a higher accuracy (57.7%) than the RF and RCM-MLC classifications, indicating that from a MCS classification perspective, a *best band* approach is a worthy undertaking. However, the *MCS agreement map* (Fig. 13a) and agreement map with areas of uncertainty masked (Fig. 13c) provided accuracies of 69.8% and 71.6%, respectively. The RF classification, when uncertainty was removed as discussed previously, also offered a high accuracy of 71.8%. However, the increasing accuracy is proportional to the degree of uncertainty. A low tolerance for uncertainty in the classification process results in fewer pixels being classified and, as one would expect, a higher overall classification accuracy.

Uncertainty Analysis

The RCM-MLC uncertainty (variability) map (Fig. 7b) produced for the data combination that provided the best classification (i.e. the highest overall accuracy) of the thirteen classes, which was the Landsat-7 (raw, ratios), was compared to the majority classification map (Fig. 7a) to determine if there was a relationship between class and areas of high classification uncertainty. The classes with the highest uncertainty (class 7 on Fig. 7b) included the sandstones and evaporites. The same comparison was made with the generalized lithological classes, in which Landsat-8 (raw, ratios) provided the highest accuracy (Fig. 9a, b). The class that experienced the highest uncertainty was, once again, sandstone. This indicates that sandstone is characterized by the most spectral diversity and was not fully represented by the associated training areas. However, as mentioned above, the sandstones were separable using the Aster TIR data, because of their unique thermal response.

DISCUSSION

This study has indicated that supervised classification techniques can play a useful role in the generation of bedrock and surficial maps of northern environments, providing that the

Table 7. Ranking of 30 of the 43 variables (image bands) constituting input to the Random Forest (RF) classification.

| RANK (group – see Fig. 11) | NUMBER in <i>gini index</i> plot (Fig. 11) | DATA | BAND |
|----------------------------------|--|-----------------------|--------------------|
| Group 1 | | | |
| 1 | 8 | L7 | 5/4 – ferrous iron |
| 2 | 18 | L8 | Band 5 – NIR |
| Group 2 | | | |
| 3 | 11 | L8 | 6/5 – ferrous iron |
| 4 | 13 | L8 | 5/4 biomass |
| Group 3 | | | |
| 5 | 34 | ASTER | Band 4 (NIR) |
| 6 | 4 | L7 | Band 4 (NIR) |
| 7 | 19 | L8 | Band 6 (SWIR) |
| 8 | 5 | L7 | Band 5 (SWIR) |
| 9 | 20 | L8 | Band 7 (SWIR) |
| 10 | 3 | L7 | Band 3 (red) |
| Group 4 | | | |
| 11 | 6 | L7 | Band7 (SWIR) |
| 12 | 2 | L7 | Band 2 (green) |
| 13 | 26 | ASTER | Band 6 (SWIR) |
| 14 | 1 | L7 | Band 1 (blue) |
| Group 5 | | | |
| 15 | 35 | ASTER | Band 5 (SWIR) |
| 16 | 12 | L8 | 6/7 - clay |
| Group 6 | | | |
| 17 | 17 | L8 | Band 4 (red) |
| 18 | 37 | ASTER | Band 7 (SWIR) |
| 19 | 38 | ASTER | Band 8 (SWIR) |
| 20 | 43 | Total Field Magnetics | |
| 21 | 7 | L7 | 5/7 -clay |
| 22 | 31 | ASTER | 4/1 - |
| 23 | 39 | ASTER | Band 9 (SWIR) |
| 24 | 29 | ASTER | 4/2 |
| Group 7 | | | |
| 25 | 28 | ASTER | 4/3 |
| 26 | 15 | L8 | Band 1b – (blue) |
| 27 | 16 | L8 | Band 3 (green) |
| 28 | 42 | ASTER | Band 3 (NIR) |
| 29 | 10 | L7 | 4/3 |
| 30 | 41 | ASTER | Band 2 (red) |

rocks and overburden demonstrate good spectral separation. These suitable environments tend to occur in Arctic Island and coastal regions where less lichen cover occurs because of favourable meso-climate conditions. The classified maps presented herein can be generated to accept more uncertainty but with large areas (i.e. more pixels) classified, producing what looks more like a traditional geologic map.

The advantage of semi-automatic classification techniques, in addition to their efficiency and speed, is that metrics that assess the uncertainty of the classification map can also be generated. However, as demonstrated here, a reduction in uncertainty results in fewer pixels being classified, albeit with greater overall confidence. These maps can be used as preliminary maps to aid in field mapping activities, as well as within a GIS environment,

where they can guide more detailed visual interpretation to fill-in unclassified pixels.

This study shows that a new MCS classification approach using *majority classification, agreement* and *best band* maps can result in better agreement with a generalized lithological map (Fig. 2b). It should be noted that the geology map used for comparison purposes was produced by using an integrated approach comprising the visual interpretation of stereo Spot-5 images and detailed field work.

Landsat-7 and Landsat-8 (raw and ratios) data combinations resulted in the highest classification accuracies using RCM-MLC. Landsat-8 offers an additional band in the blue portion of the electromagnetic spectrum and wavelengths similar to Landsat-7, except that the bandwidths are narrower. Both images were acquired in July, which provides optimum conditions because of higher sun elevation angles. The higher radiometric resolution of Landsat-8 (16-bit) did not offer any significant advantages with respect to classification accuracy compared to Landsat-7 (8-bit). It is comforting to confirm the good performance of Landsat-8, which will provide uninterrupted world coverage once Landsat-7, the workhorse of remote sensing applications for many years, expires.

The poor performance of Aster is no doubt related to image acquisition in mid September, when the sun angle was only 20°, resulting in a lower signal-to-noise ratio and an increased topographic effect such as shadowing and low reflectance from backslopes. The two Aster scenes were corrected and mosaicked, and although care was taken to radiometrically balance the scenes, these factors may also negatively affect the classification results. As such, the comparison with Landsat may not be made on a level playing field; i.e. better results may have been obtained from the Aster data if acquisition had occurred in July. The VNIR and SWIR Aster bands provided better results than the TIR bands, except for sandstone, the thermal properties of which were much different than the surrounding rocks. None of the TIR bands were selected by the RF classifier as being strongly predictive (Table 7). Thermal

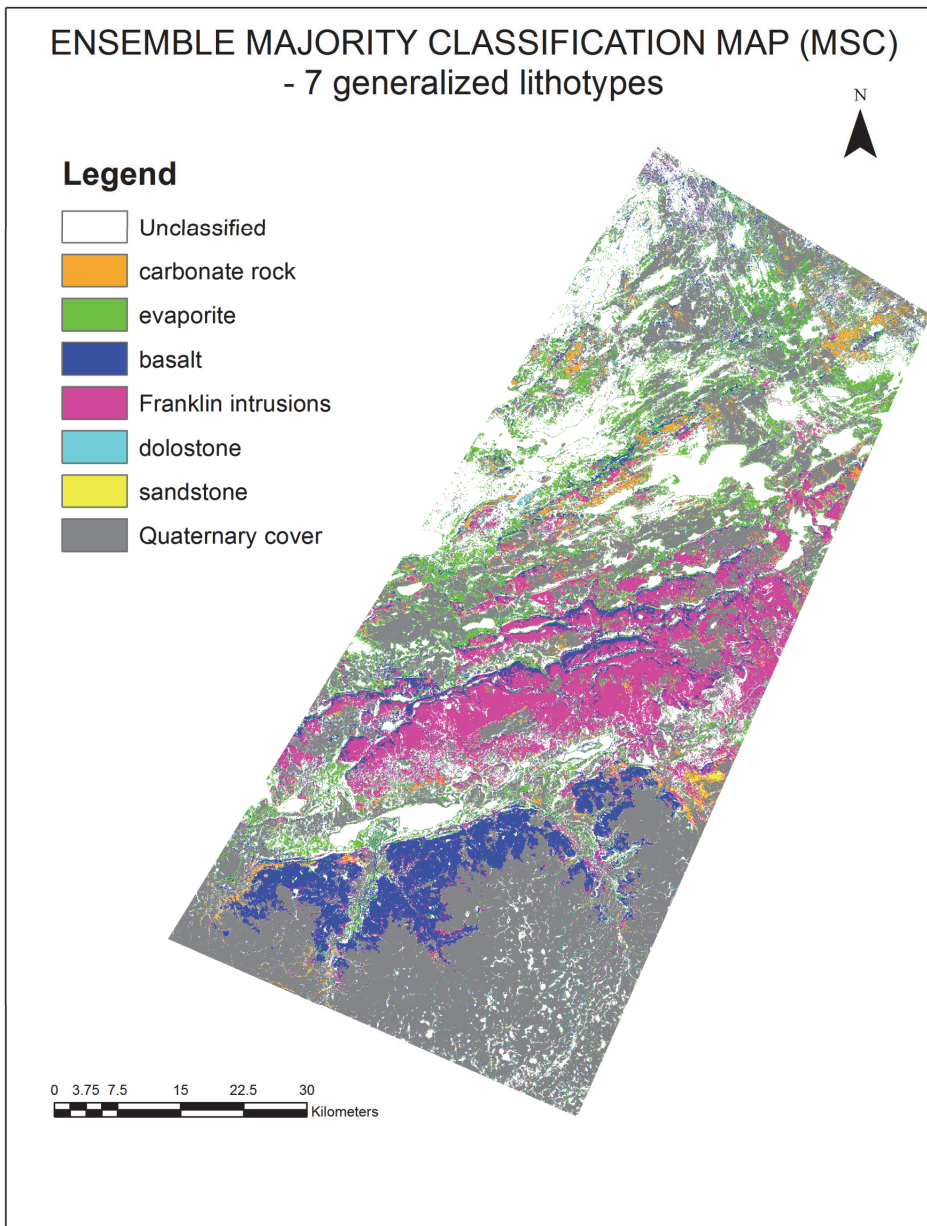


Figure 12. MLC Multiple Classification System (MCS) Majority Classification Map.

data are better collected during the night to suppress solar heating effects; however, northern environments, although dark for six months of the year, are completely snow-covered during these winter months, suppressing thermal signatures from the underlying rocks.

The Spot-5 data also provided lower classification results than the Landsat data. This may also be related to the lower sun angle on the date of acquisition (22 August), and also because of fewer spectral bands. Behnia et al. (2012), in another remote study of Victoria Island, note that Landsat-7 data using RCM-MLC pro-

vided a higher classification accuracy than Spot-5. They attributed this result to the higher spectral resolution of Landsat, which records reflectance in the blue and SWIR portions of the electromagnetic spectrum. They also concluded that the lower spatial resolution of Landsat actually acts as a natural filter by generalizing the surface and presenting less spectral variability than the higher spatial resolution of Spot-5 data.

The addition of ratios for all sensors (Landsat-7, Landsat-8, and Aster) improved overall classification accuracies for both the (thirteen) detailed and (seven) generalized litho-

logical classes. Ratios are advantageous to the classification procedure as they help to reduce illumination effects related to topography and, more importantly, emphasize spectral differences associated with individual minerals and groups of minerals that help to identify various lithotypes. In fact, the best predictor, established through the RF classifier, was a Landsat-7 band 5/4 ratio (Table 7) designed to detect rocks with a high ferrous iron content.

Inclusion of magnetic data (total field or converted to magnetic susceptibility) in the classification process is also recommended, as it can help to identify rocks characterized by either high or low magnetic susceptibilities, as demonstrated in this study and by Martel et al. (2005), Harris et al. (2008b, c), and Schetselaar and Ryan (2008). Gamma ray spectrometry data, although not used in this study, should also be included in the classification process as it records information on the content of uranium, thorium and potassium in rocks at the earth's surface. These elements are present in most rock-forming minerals: potassium is a common constituent in feldspar, biotite, and muscovite, whereas uranium and thorium generally occur in low concentrations in a wide range of accessory minerals. Unfortunately, no gamma ray data were available for the study area; however, Harris et al. (2009) and Ford et al. (2008) have shown that these data are extremely useful in the classification of bedrock types.

The selection of representative training areas is obviously a key factor in the application of machine learning algorithms such as classification to the mapping of bedrock or surficial materials. Even within classes, spectral variability may be present. Ensemble classification algorithms such as RCM-MLC and RF are used to account for variability in training area signatures by employing *bagging* and cross-validation methods (see Harris et al. 2012b). This can assist in identifying training areas that can be modified and/or deleted from the classification process. Furthermore, these classification algorithms provide a more robust estimate of classification accuracies and generate uncertainty metrics to accompany the classified map.

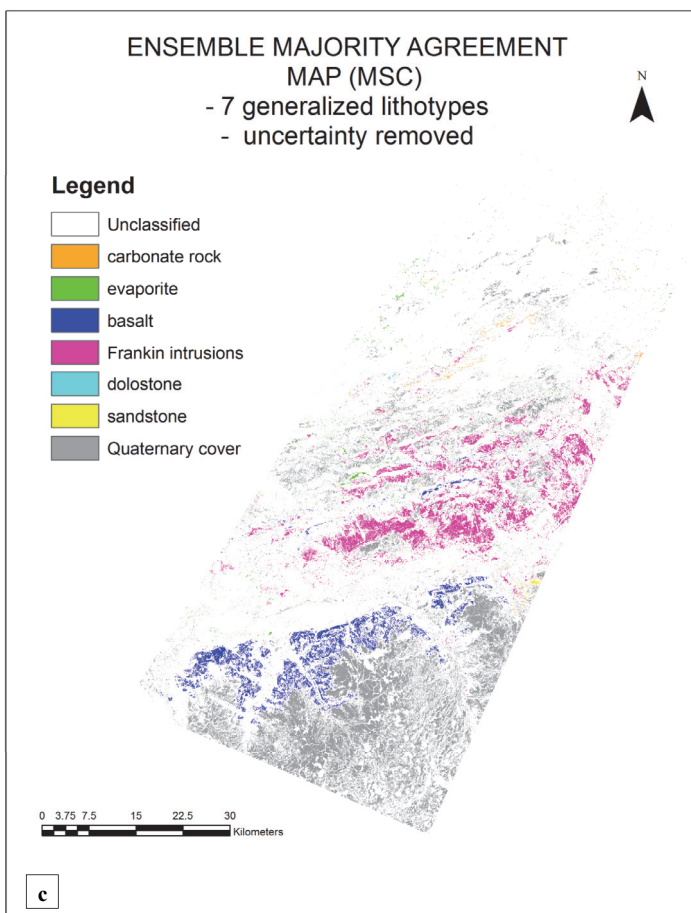
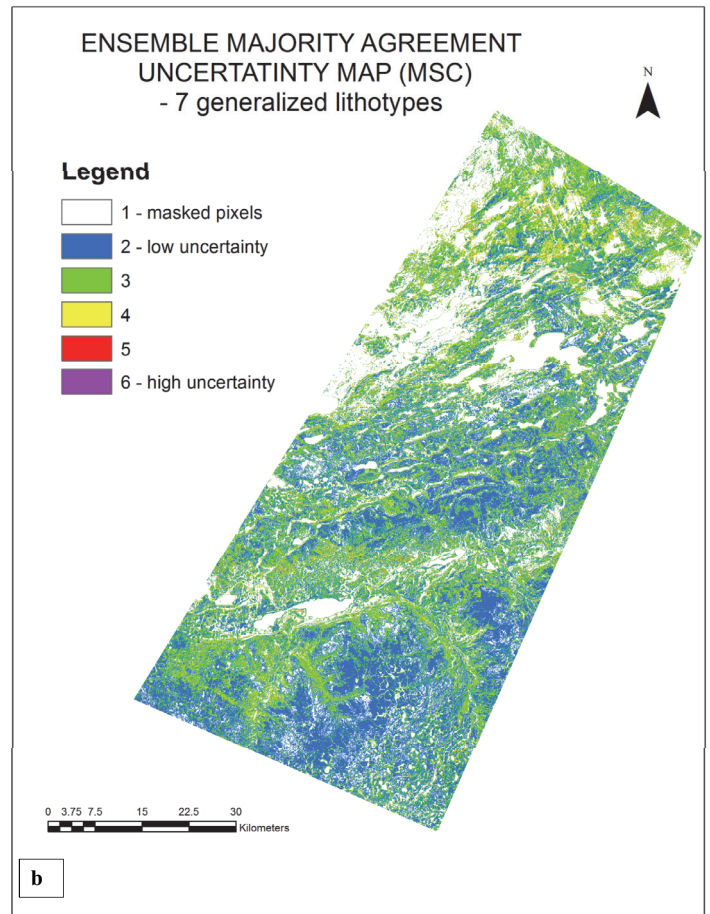
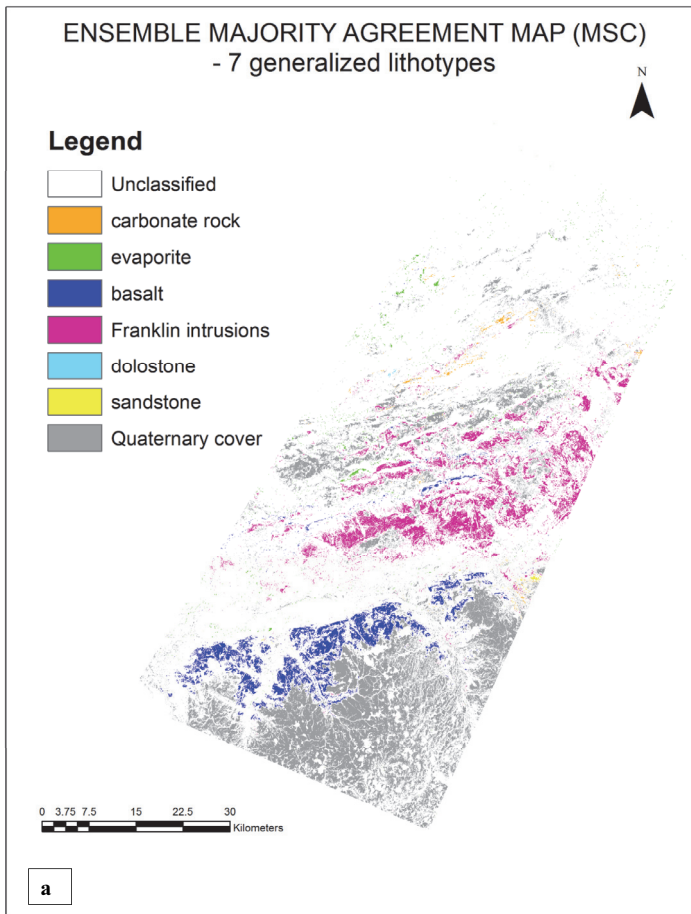


Figure 13. (a) Multiple Classification System (MCS) Classification Agreement Map. (b) MCS Agreement Uncertainty Map. (c) MCS Classification Agreement Map with uncertainty (class 5 and 6) from the ensemble uncertainty map shown in b, removed.

Table 8. Best band data combination for each of the seven generalized lithological – surficial classes.

| Class | Data combination | Class accuracy (%) |
|------------|--|--------------------|
| Carbonates | LANDSAT 8 –raw bands | 96.2 |
| Evaporites | LANDSAT 7 –raw, ratios, mag | 98.6 |
| Basalt | ASTER – raw, ratios LANDSAT 7, raw, ratios, mag | 90.9 90.7 |
| Diabase | LANDSAT 7 – raw ,ratios, mag | 90.5 |
| Dolostone | LANDSAT 8 – raw, ratios | 96.8 |
| Sandstone | ASTER- raw, ratios | 86.7 |
| Quaternary | LANDSAT 8 – raw, ratios | 94.9 |

Comparing the RCM–MLC and RF algorithms is difficult because different datasets were used: for RF, all 43 bands were employed, whereas RCM–MLC used various subsets of the 43 bands. Use of MLC with a restricted number of training areas would result in the *Hughes effect* (Hughes 1968; Oommen et al. 2008; Alonso et al. 2011), which describes how classification accuracy depends on the number of training samples. In general, the separability of classes (as well as the number of statistical parameters defining the classes) increases with the number of dimensions (bands). In classification, a fixed number of training samples, generally proportional to the number of bands to be classified, are used, and at some point the classification accuracy can decrease if the number of training samples is insufficient. High dimensional data are generally more difficult to work with, as many bands can increase the noise factor and hence error factor, requiring many training samples to be used. Story and Congalton (1986) suggested that at least 50 pixels per map class should be used in the training process, whereas Jensen (2005) suggested $10n$ pixels for each class, where n equals the number of bands (channels) used in the classification. Other researchers suggest 10 to $30n$ pixels per class (Mather 1999; Piper 1999). Inclusion of the total number of input bands (43) for the RCM–MLC classification could lead to an overestimation of accuracy because of the limited number of training areas used in this study (see Table 1). However, in using RCM–MLC, the *Hughes effect* can be reduced because the iterative random sampling of the training dataset (*bagging*) produces a series of classifications as well as independent cross-validations (see above and Harris et al. 2012b). RF is also less affected by the *Hughes effect* because the *bagging* process is used to create, in this case, 100 trees (classifications) and an equal number of validation estimates. Furthermore, RF has been shown, in a number of studies, to outperform other classifiers (Gislason et al. 2006; Crisci et al. 2012; Craknell and Reading 2014). In addition, RF provides a predictive rank for each input band and is a non-parametric classifier. In this study, the highest

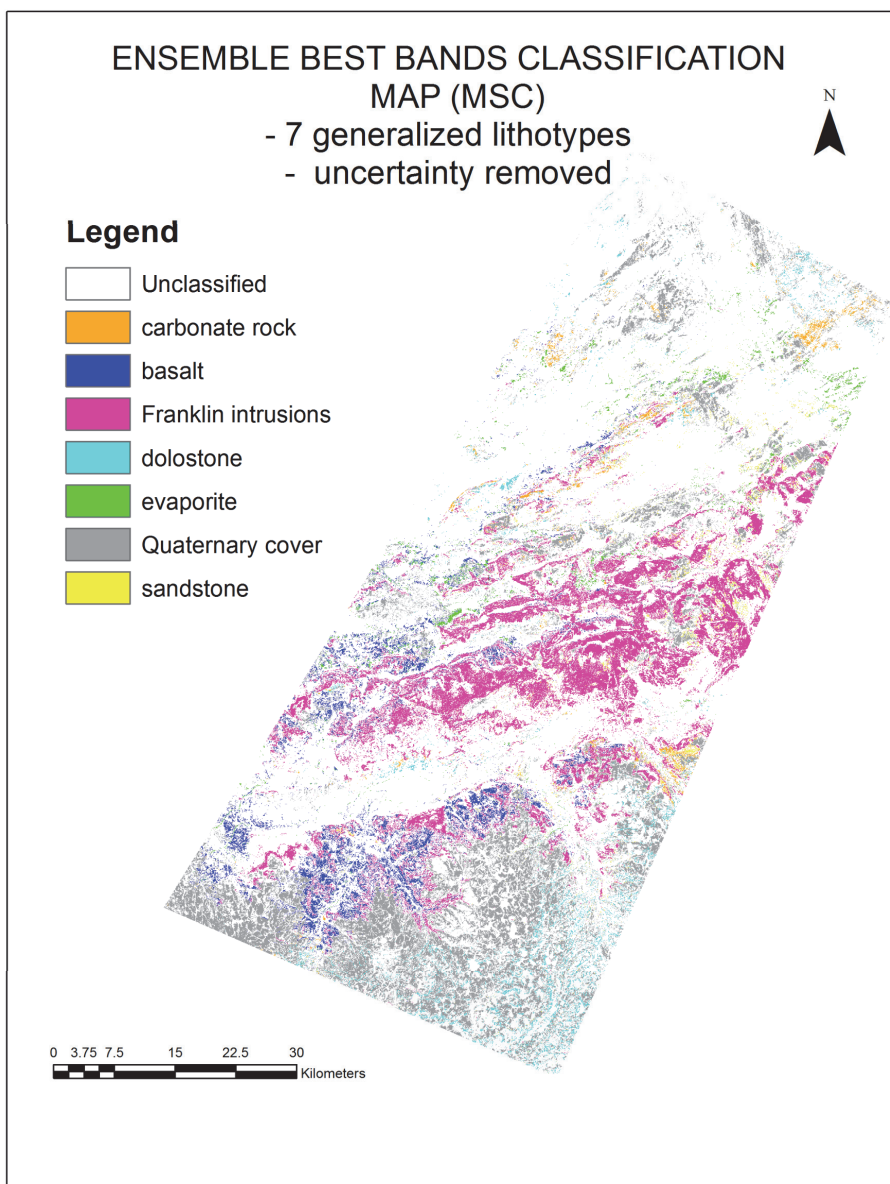


Figure 14. Multiple Classification System (MCS) Best Bands Classification Map using thresholded average rule images (>95%) to reduce uncertainty (see text).

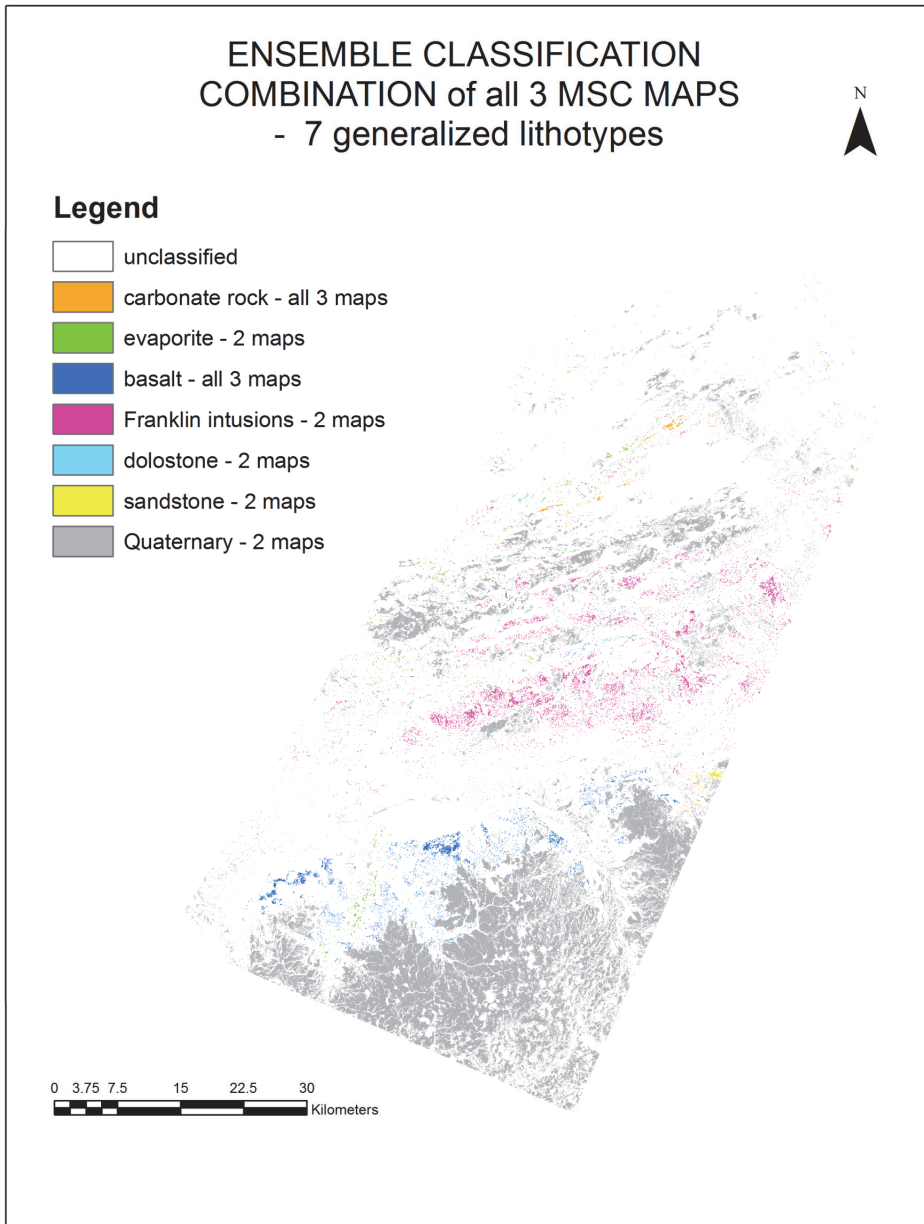


Figure 15. The Multiple Classification System (MCS) Combination Map combines MCS agreement (Fig. 13a), best bands (Fig. 14), and majority classification (Fig. 12) maps.

overall classification accuracy (99%) was achieved using RF, bearing in mind that all 43 bands (raw and ratios) derived from the Landsat, Aster and Spot data were used in the classification experiments.

In northern terranes where remotely sensed data are often acquired at lower sun angles, such as the Aster data used herein, it is recommended to either topographically correct remotely sensed data or, in the present case, to account for illumination effects associated with topography. This was accomplished simply by

incorporating a 1:50,000 DEM (Canadian Digital Elevation Data) in which hillshade maps using the sun azimuth and elevation values were determined from the acquisition parameters of each sensor. The hillshade maps were then thresholded to mask areas in shadow and backslope areas that reflect less energy, both of which can add an element of confusion to the classification process.

Comparison of a classified map derived from even moderate-resolution remotely sensed data with a traditional geologic map is problematic as

the two are constructed using entirely different processes. Although the geology map used here for comparison to the classified maps did incorporate a component of remote sensing (visual interpretation of stereo Spot-5 imagery), the map was also based on detailed field work. Furthermore, the cognitive processes involved in traditional map-making, which involves the incorporation of process information and generalization based on field experience, is more difficult to capture by machine-learning algorithms. Remote sensors collect data on a grid cell basis (30m for Landsat), in which much spectral heterogeneity is recorded. A traditional geological map is often a generalized model of a geological environment in that it presents more homogeneous classes. The fact that remotely sensed data records more heterogeneous information does not render the classified map useless; in fact, just the opposite occurs, as indicated in this study. The more detailed classification (thirteen classes) produced a map with spectral sub-units within the main lithological groups. To create a level playing field for the comparison, we generalized both the geological map and the training areas to seven basic lithological – surficial classes. The majority of the classification maps showed some agreement (40 to 50%) with the generalized geology map. However, an ensemble MCS classification approach, here employing *majority agreement* and *best band* maps, resulted in higher classification accuracies in excess of 70% when compared to the geology map. The trade-off is that fewer pixels are classified, but with more confidence. Furthermore, inclusion of uncertainty metrics (i.e. RCM-MLC variability, RF probability and ensemble range maps) increases overall accuracies, again at the expense of the number of pixels classified. The geologist has the option of using classification maps in which all pixels are classified other than those that have been purposely masked, or else using a map that has had fewer pixels classified, but with less spectral uncertainty.

CONCLUSION

This study has demonstrated that classification of remotely sensed optical data can significantly contribute to the

Table 9. Summary of the overall classification accuracy for various data combinations compared to the generalized geology map (Figure 2b).

| Data | Accuracy | Kappa |
|---|----------|-------------------|
| LANDSAT – raw bands | 43.9 | 0.27 |
| ASTER – raw bands and ratios | 34.2 | 0.16 |
| LANDSAT 7 – raw, ratios, magnetics | 44.9 (1) | 0.27 |
| LANDSAT 7 – raw bands and ratios | 43.9 | 0.27 |
| LANDSAT 8 – raw bands | 39.4 | 0.24 |
| LANDSAT 8 – raw bands and ratios | 42.4 | .25 (Fig. 9a) |
| LANDSAT 8 – raw, ratios, mag | 46.1 (1) | 0.27 |
| ENSEMBLE agreement map | 69.8 | 0.557 (Fig. 13a) |
| ENSEMBLE agreement map with uncertainty removed | 71.6 | 0.59 12 (Fig.13c) |
| ENSEMBLE RCM majority classification map | 48.4 | .32 (Fig. 12) |
| Ensemble best bands RCM classification | 57.7 | 0.37 (Fig. 14) |
| RF classification | 42.4 | 0.25 (Fig. 10a) |
| RF classification with uncertainty removed | 71.8 | 0.60 (Fig. 10b) |

geologic mapping process in a geological environment in which the spectral expression of rocks is optimal. These environments tend to occur in Arctic islands and coastal environments in Canada’s North. A map in which all non-masked pixels are classified or a map with fewer pixels classified, but with more spectrally certainty, can be produced. The maps can provide first-order geologic information in poorly-mapped or unmapped areas, as well as preliminary maps to support and focus field-based studies. For example, a classified map produced from an ensemble classification in which very uncertain areas have been removed (e.g. Fig. 13c) could form the basis of a more detailed visual analysis of various remotely sensed data (2D and 3D) in a GIS environment using a *heads-up* interpretation process and touch-screen technology (see Harris et al. 2012a).

Landsat-8 data produced classification results similar to those of Landsat-7, which will ensure the reliable and continuous use of the Landsat series of satellites for geological mapping into the future. The performance of Aster data was limited by the season of acquisition (low sun angle), indicating that for Arctic research, mid-June to mid-August data acquisition is best for optimal results. It is expected

that Aster would have been more useful if a July acquisition date could have been found. Certainly, the Aster imagery did classify certain lithotypes more accurately than did Landsat-7 or Landsat-8, although the overall performance was inferior. Aster TIR data may see only limited applications in northern environments because of snow cover and day-time solar heating biases. Spot-5 data provides higher spatial resolution, which is very useful for detailed visual interpretation, but has lower spectral resolution, lacking reflectance information in the blue and SWIR wavelengths.

For geological classification, both raw optical data and calculated ratios should be used to directly focus on spectral differences related to mineralogy, primarily in iron-, clay-, and carbonate-bearing rocks. Magnetic and gamma ray spectrometry data, depending on the geological environment, should also be incorporated.

Both RCM-MLC and RF produced high classification accuracies. The RF algorithm may be preferred, as other studies indicate that it marginally outperforms other classification algorithms by 2–10% with respect to classification accuracies. However, more research is required to confirm this. Furthermore, RF does not require nor-

mal distribution of the training data to be classified, provides a ranked list of the best predictors, is less sensitive to the *Hughes effect*, and does not overfit the data.

A new MCS approach, using a *majority classification agreement* and *best bands* map derived from different remotely sensed data types, is recommended for other geologic classification studies in northern environments. This requires more data acquisition, but as shown in this study, results in higher classification accuracies when compared to the new geology map. This MCS classification approach could be useful for surficial mapping applications using multiple image acquisitions (time series) by the same or different sensors over the northern summer (mid-June to end of July). This time-series approach may prove useful for capturing biomass and soil moisture differences that may be helpful in the mapping of surficial materials.

ACKNOWLEDGEMENTS

The authors would like to thank the Geological Survey of Canada under the Remote Predictive Mapping Project (RPM), part of the Geo-mapping for Energy and Minerals (GEM) program, for supporting and funding this research. We would also like to thank two anonymous reviewers whose suggestions and comments greatly improved this manuscript. This is Geological Survey of Canada contribution number 20130439.

REFERENCES

Abrams, M., and Hook, S., 1995, Aster Users Handbook: Jet Propulsion Laboratory, Pasadena, CA, 133 p. Available from: http://gds.aster.ersdac.jspacesystems.or.jp/gds_www2000/service_e/u.tools_e/cross/GUIDEED.PDF

Alonso, M.C., Malpica, J.A., and Martinez de Aguirre, A., 2011, Consequences of the Hughes Phenomenon on some classification techniques: American Society for Photogrammetry and Remote Sensing 2011 Annual Conference, Milwaukee, WI, May 1-5th.

Banfield, R.E., Hall, L.O., Bowyer, K.W., and Kegelmeyer, W.P., 2007, A comparison of decision tree ensemble creation techniques: IEEE Transactions on Pattern Analysis and Machine

- Intelligence, v. 29, p. 173–180, <http://dx.doi.org/10.1109/TPAMI.2007.250609>.
- Behnia, P., Harris, J.R., Rainbird, R.H., Williamson, M.C., and Sheshpari, M., 2012, Remote predictive mapping of bedrock geology using image classification of Landsat and SPOT data, western Minto Inlier, Victoria Island, Northwest Territories, Canada: *International Journal of Remote Sensing*, v. 33, p. 6876–6903, <http://dx.doi.org/10.1080/01431161.2012.693219>.
- Benediktsson, J.A., and Swain, P.H., 1992, Consensus theoretic classification methods: *IEEE transactions on Systems, Man, and Cybernetics*, v. 22, p. 688–704, <http://dx.doi.org/10.1109/21.156582>.
- Benediktsson, J.A., Chanussot, J., and Fauvel, M., 2007, Multiple classifier systems in remote sensing: from basic to recent developments, *in* Kittler, H.J., and Roli, F., *eds.*, *Multiple Classifier Systems: Proceedings of the 7th International Workshop, Prague, Lecture notes in Computer Science 4472*, Springer, Berlin.
- Biggar, S.F., Thome, K., McCorkel, J.T., and D'Amico, J.M., 2005, Vicarious calibration of the ASTER sensor including crosstalk correction, *in* Butler, J.J., *ed.*, *Earth Observing Systems X Earth Observing Systems X: Proceedings of the SPIE*, v. 5882, p. 408–415.
- Breiman, L., 1996, Bagging predictors: *Machine Learning*, v. 24, p. 123–140, <http://dx.doi.org/10.1007/BF00058655>.
- Breiman, L., 2001, Random Forests: *Machine learning*, v. 45, p. 5–32, <http://dx.doi.org/10.1023/A:1010933404324>.
- Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., 1984, *Classification and regression trees: Wadsworth and Brooks, Cole Statistics/Probability Series*, Pacific Grove, CA.
- Chuvieco, E., and Huete, A., 2010, *Fundamentals of Satellite Remote Sensing: CRC Press, Boca Raton, FL*, 235 p.
- Craknell, M.J., and Reading, A.M., 2014, Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information: *Computers and Geosciences*, v. 63, p. 22–33, <http://dx.doi.org/10.1016/j.cageo.2013.10.008>.
- Crisci, C., Ghattas, B., and Perera, G., 2012, A review of supervised machine learning algorithms and their applications to ecological data: *Ecological Modelling*, v. 240, p. 113–122, <http://dx.doi.org/10.1016/j.ecolmodel.2012.03.001>.
- Desnoyers, D.W., and Harris, J.R., 2003, Remote Predictive Mapping - A new tool for geoscience mapping (abstract): 31st Yellowknife Geoscience Forum, 19–21 November, Abstract.
- Doan, H.T.X., and Foody, G.M., 2007, Increasing soft classification accuracy through the use of an ensemble of classifiers: *International Journal of Remote Sensing*, v. 28, p. 4609–4623, <http://dx.doi.org/10.1080/01431160701244872>.
- Drury, S., 1993, *Image Interpretation in Geology*, 2nd Edition: Chapman and Hall, London, UK, 173 p.
- El Rakaiby, M.L., 1995, The use of enhanced Landsat-TM image in the characterization of uraniferous granitic rocks in the Central Eastern Desert of Egypt: *International Journal of Remote Sensing*, v. 16, p. 1063–1074, <http://dx.doi.org/10.1080/01431169508954463>.
- Ford K., Harris, J.R., Shives, R., Carson, J., and Buckle, J., 2008, Gamma Ray Spectrometry, *in* Harris, J.R., *ed.*, *Remote Predictive Mapping: an aid for northern mapping: Geological Survey of Canada, Open File 5643*, p. 53–97.
- Freund, Y., and Schapire, R.E., 1999, A short introduction to boosting: *Journal of Japanese Society for Artificial Intelligence*, v. 14, p. 771–780.
- Gislason, P.O., Benediktsson, J.A., and Sveinsson, J.R., 2006, Random Forests for land cover classification: *Pattern Recognition Letters*, v. 27, p. 294–300, <http://dx.doi.org/10.1016/j.patrec.2005.08.011>.
- Glikson, A.Y., and Creasey, J.W., 1995, Application of Landsat-5 TM imagery to mapping of the Giles Complex and associated granulites, Tomkinson Range, western Musgrave Block, central Australia: *AGSO Journal of Australian Geology and Geophysics*, v. 16, p. 173–193.
- Gupta, R.P., 1991, *Remote Sensing Geology: Springer-Verlag, Berlin*, 356 p., <http://dx.doi.org/10.1007/978-3-662-12914-2>.
- Harris, J.R., *ed.*, 2008, *Remote Predictive Mapping: an aid for northern mapping: Geological Survey of Canada, Open File 5643*, 306 p., 1 DVD, <http://dx.doi.org/10.4095/225997>.
- Harris, J.R., Rogge, D., Hitchcock, R., Ijewliw, O., and Wright, D., 2005, Mapping lithology in Canada's Arctic: application of hyperspectral data using the minimum noise fraction transformation and matched filtering: *Canadian Journal of Earth Sciences*, v. 42, p. 2173–2193, <http://dx.doi.org/10.1139/e05-064>.
- Harris, J.R., Schetselaar, E.M., Lynds, T., and de Kemp, E.A., 2008a, Remote predictive mapping: A strategy for geological mapping of Canada's North, Chapter 2, *in* Harris, J.R., *ed.*, *Remote Predictive Mapping: an aid for northern mapping: Geological Survey of Canada, Open File 5643*, p. 5–27, <http://dx.doi.org/10.4095/225997>.
- Harris J.R., Schetselaar, E., de Kemp, E.A., and St-Onge, M., 2008b, LANDSAT, Magnetic and Topographic Data for Regional Lithological Mapping – Southeast Baffin Island, Case study, *in* Harris, J.R., *ed.*, *Remote Predictive Mapping: an aid for northern mapping: Geological Survey of Canada, Open File 5643*, p. 173–182, <http://dx.doi.org/10.4095/225997>.
- Harris, J.R., Martel, E., Currie, M., Pierce, K., Pilkington, M., and Keating, P., 2008c, Snowbird Lake –NWT: Application of RPM methods, *in* Harris, J.R., *ed.*, *Remote Predictive Mapping: an aid for northern mapping: Geological Survey of Canada, Open File 5643*, p. 161–171, <http://dx.doi.org/10.4095/225997>.
- Harris, J.R., Ford, K.L., and Charbonneau, B.W., 2009, Application of gamma ray spectrometer data for lithological mapping in a cordilleran environment, Sekwi region, NWT: Special Issue of the *Canadian Journal of Remote Sensing*, v. 35, p. S12–S30, <http://dx.doi.org/10.5589/m09-022>.
- Harris, J.R., McGregor, R., and Budkewitsch, P., 2010, Geological analysis of hyperspectral data over southwest Baffin Island: methods for producing spectral maps that relate to variations in surface lithologies: *Canadian Journal of Remote Sensing*, v. 36, p. 412–435, <http://dx.doi.org/10.5589/m10-072>.
- Harris, J.R., Schetselaar, E., and Behnia, P., 2012a, Remote Predictive Mapping: An Approach for the Geological Mapping of Canada's Arctic, *in* Dar, I.A., *ed.*, *Earth Sciences: InTech*, p. 495–524. Available from: <http://cdn.intechopen.com/pdfs-wm/27604.pdf>.
- Harris, J.R., Grunsky, E.C., He, J., Gorodetzky, D., and Brown, N., 2012b, A Robust, cross-validation classification method (RCM) for improved mapping accuracy and confidence metrics: *Canadian Journal of Remote Sensing*, v. 38, p. 69–90,

- <http://dx.doi.org/10.5589/m12-013>.
- Hughes, G.F., 1968, On the mean accuracy of statistical pattern recognizers: *IEEE Transactions on Information Theory*, v. 14, p. 55–63, <http://dx.doi.org/10.1109/TIT.1968.1054102>.
- Hulbert, L.J., Rainbird, R.H., Jefferson, C.W., and Friske, P., 2005, Map of mafic and ultramafic bodies related to the Franklin magmatic event, Minto Inlier, Victoria Island: Geological Survey of Canada, Open File No. 4928, <http://dx.doi.org/10.4095/220616>.
- Jensen, J.R., 2005, Introductory digital image processing: a remote sensing perspective: Pearson Prentice Hall, 3rd ed., Upper Saddle River, NJ.
- Jones, H.G., and Vaughan, R.A., 2010, Remote sensing of vegetation: Principles, Techniques, and Applications: Oxford University Press, 384 p.
- Kalinowski, A., and Oliver, S., *compilers*, 2004, ASTER mineral index processing manual: Remote Sensing Applications, Geoscience Australia, internal report 36, 37 p., http://www.ga.gov.au/image_cache/GA7833.pdf.
- Kavak, K.S., 2005, Determination of palaeotectonic and neotectonic features around the Menderes Massif and the Gediz Graben (western Turkey) using Landsat TM image: *International Journal of Remote Sensing*, v. 26, p. 59–78, <http://dx.doi.org/10.1080/01431160410001709994>.
- Leverington, D.W., 2001, Discriminating lithology in Arctic environments from Earth orbit: an evaluation of satellite imagery and classification algorithms: Unpublished PhD thesis, Department of Geological Sciences, University of Manitoba, MB.
- Leverington, D.W., 2010, Discrimination of sedimentary lithologies using Hyperion and Landsat Thematic Mapper data: a case study at Melville Island, Canadian High Arctic: *International Journal of Remote Sensing*, v. 31, p. 233–260, <http://dx.doi.org/10.1080/01431160902882637>.
- Leverington, D.W., and Moon, W.M., 2012, Landsat-TM based discrimination of lithological units associated with the Purtuniqu Ophiolite, Quebec, Canada: *Remote Sensing*, v. 4, p. 1208–1231, <http://dx.doi.org/10.3390/rs4051208>.
- Lillesand, M.L., Kiefer, R.W., and Chipman, J.W., 2004, Introduction to Remote Sensing: John Wiley and Sons, Canada, 763 p.
- Lorenz, H., 2004, Integration of Corona and Landsat Thematic Mapper data for bedrock geological studies in the high Arctic: *International Journal of Remote Sensing*, v. 25, p. 5143–5162, <http://dx.doi.org/10.1080/01431160410001705097>.
- Macias, L.F., 1995, Remote sensing of mafic-ultramafic rocks: Examples from Australian Precambrian terranes: *AGSO Journal Australian Geology and Geophysics*, v. 16, p. 163–171.
- Martel, E., Harris, J.R., Currie, M., and Pierce, K., 2005, Snowbird Lake (NTS 65D) Remote Predictive Mapping and Geoscience Data compilation: Northwest Territories, Open File 2005-08, 1 CD-ROM.
- Mather, P.M., 1999, Computer Processing of Remotely Sensed Images, 2nd Edition: Chichester, John Wiley and Sons, 352 p.
- Menze, B.H., Kelm, B.M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., and Hamprecht, F.A., 2009, A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data: *BMC Bioinformatics*, v. 10, 213, <http://dx.doi.org/10.1186/1471-2105-10-213>.
- Oommen, T., Misra, D., Twarakavi, N.K.C., Prakash, A., Sahoo, B., and Bandopadhyay, S., 2008, An Objective Analysis of Support Vector Machine Based Classification for Remote Sensing: *Mathematical Geoscience*, v. 40, p. 409–424, <http://dx.doi.org/10.1007/s11004-008-9156-6>.
- Peña, S.A., and Abdelsalam, M.G., 2006, Orbital remote sensing for geological mapping in southern Tunisia: Implication for oil and gas exploration: *Journal of African Earth Science*, v. 44, p. 203–219, <http://dx.doi.org/10.1016/j.jafrearsci.2005.10.011>.
- Piper, J., 1999, The effect of zero feature correlation assumption on maximum likelihood based classification of chromosomes: *Signal processing*, v. 12, p. 49–57.
- Polikar, R., 2006, Ensemble based systems in decision making: *IEEE Circuits and Systems Magazine*, v. 6, p.21–45, <http://dx.doi.org/10.1109/MCAS.2006.1688199>.
- Rainbird, R.H., Jefferson, C.W., Hildebrand, R.S., and Worth, J.K., 1994, The Shaler Supergroup and revision of Neoproterozoic stratigraphy in Amundsen, Northwest Territories: Geological Survey of Canada, Current Research no. 1994-C, p. 61–70.
- Rainbird, R.H., Jefferson, C.W., and Young, G.M., 1996, The early Neoproterozoic sedimentary Succession B of northwestern Laurentia: Correlations and paleogeographic significance: *Geological Society of America Bulletin*, v. 108, p. 454–470, [http://dx.doi.org/10.1130/0016-7606\(1996\)108<0454:TENSSB>2.3.CO;2](http://dx.doi.org/10.1130/0016-7606(1996)108<0454:TENSSB>2.3.CO;2).
- Rainbird, R.H., Harrison, J.C., Hillary, E.M., Ford, A., Hulbert, L.J., Christie, R.L., and Campbell, F.H.A., 2013a, Geology, tectonic assemblage map of Hadley Bay, Victoria and Prince of Wales islands, Nunavut - Northwest Territories: Geological Survey of Canada, Canadian Geoscience Map 75, (*edition preliminary*), scale: 1:500,000, 1 sheet, <http://dx.doi.org/10.4095/292820>.
- Rainbird, R.H., Bédard, J.H., Dewing, K., Hadlari, T., and Thomson, D., 2013b, Geology, Qiqittivik, Victoria Island, Northwest Territories: Geological Survey of Canada, Canadian Geoscience Map 59, (*edition preliminary*), scale: 1:500,000, 1 sheet, <http://dx.doi.org/10.4095/292820>.
- Rainbird, R.H., Christie, R.L., Harrison, J.C., and Ford, A., 2013c, Geology, tectonic assemblage map of the Ulukhaktok area, southwestern Victoria Island, Nunavut - Northwest Territories: Geological Survey of Canada, Canadian Geoscience Map 77, (*edition preliminary*), scale: 1:500,000, 1 sheet, <http://dx.doi.org/10.4095/292812>.
- Rainbird, R.H., Bédard, J.H., and Williamson, N., 2013d, Geology, Takiyuaqattak, Victoria Island, Northwest Territories: Geological Survey of Canada, Canadian Geoscience Map 104, (*edition preliminary*), scale: 1:500,000, 1 sheet, <http://dx.doi.org/10.4095/293344>.
- Rajendran, S., al-Khribash, S., Pracejus, B., Nasir, S., Al-Abri, A.H., Kusky, T.M., and Ghulam, A., 2012, ASTER detection of chromite bearing mineralized zones in Semail Ophiolite Massifs of the northern Oman Mountains: *Exploration strategy: Ore Geology Reviews*, v. 44, p. 121–135, <http://dx.doi.org/10.1016/j.oregeorev.2011.09.010>.
- Richards, J.A., and Jia, X., 2006, Remote Sensing Digital Image Analysis: An Introduction: Springer-Verlag, New York, 4th Edition, 439 p.
- Rodriguez-Galiano, V.F., Chica-Olmo, M., and Chica-Rivas, M., 2014, Predictive modelling of gold potential with the integration of multisource information based on random forest: a case study on the Rodalquilar area, Southern Spain: *International Journal of Geographical Information Science*, v. 28, p. 1336–1354, <http://dx.doi.org/10.1080/13658816.2014.885527>.

- Rowan, L.C., Anton-Pacheco, C., Brickey, D.W., Kingston, M.J., Payas A., Vergo, N., and Crowley, J.K., 1987, Digital classification of contact metamorphic rocks in Extremadura, Spain, using Landsat thematic mapper data: *Geophysics*, v. 52, p. 885–897, <http://dx.doi.org/10.1190/1.1442359>.
- Rowan, L.C., Mars, J.C., and Simpson, C.J., 2005, Lithologic mapping of the Mor-dor, NT, Australia ultramafic complex by using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER): *Remote Sensing of Environment*, v. 99, p. 105–126, <http://dx.doi.org/10.1016/j.rse.2004.11.021>.
- Saadi, N.M., and Watanabe, K., 2009, Assessing image processing techniques for geological mapping: a case study in Eljufra, Libya: *Geocarto International*, v. 24, p. 241–253, <http://dx.doi.org/10.1080/10106040802556199>.
- Schetselaar, E.M., and Ryan, J., 2008, Remote Predictive Mapping Case study of the Boothia Mainland area, Nunavut, Canada, *in* Harris, J.R., *ed.*, *Remote Predictive Mapping: an aid for northern mapping*: Geological Survey of Canada, Open File 5643, p. 261–280.
- Schetselaar, E.M., Harris, J.R., Lynds, T., and de Kemp, E.A., 2007, Remote Predictive Mapping (RPM): A Strategy for Geological Mapping of Canada's North: *Geoscience Canada*, v. 34, p. 93–111.
- Story, M., and Congalton, R., 1986, Accuracy Assessment: a user's perspective: *Photogrammetric Engineering and Remote Sensing*, v. 52, p. 397–399.
- Thomson, D., Rainbird, R.H., and Dix, G., 2014, Architecture of a Neoproterozoic intracratonic carbonate ramp succession: Wynniatt Formation, Amundsen Basin, Arctic Canada: *Sedimentary Geology*, v. 299, p. 119–138, <http://dx.doi.org/10.1016/j.sed-geo.2013.11.005>.
- Thorsteinsson, R., and Tozer, E.T., 1962, Banks, Victoria and Stefansson Islands, Arctic Archipelago: Geological Survey of Canada, Memoir 330, p. 1–85.
- Van der Meer, F., Van Dijk, P.M., and Westerhof, A.B., 1995, Digital classification of the contact metamorphic aureole along the Los Pedroches batholith, south-central Spain, using Landsat Thematic Mapper data: *International Journal of Remote Sensing*, v. 16, p. 1043–1062, <http://dx.doi.org/10.1080/01431169508954462>.
- Waske, B., and Braun, M., 2009, Classifier ensembles for land cover mapping using multitemporal SAR imagery: *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 64, p. 450–457, <http://dx.doi.org/10.1016/j.isprsjprs.2009.01.003>.
- Wickert, L.M., Percival, J.B., Morris, W.A., and Harris, J.R., 2008, XRD and Infrared Spectroscopic Validation of Weathering Surfaces from Ultramafic and Mafic Lithologies Examined Using Hyperspectral Imagery, Cross Lake Area, Cape Smith Belt, Northern Quebec, Canada, *in* *Proceedings of 2008 IEEE International Geoscience and Remote Sensing Symposium*, Boston, MA, USA, 6–11 July 3, p. 362–365.

Received February 2014

Accepted as revised July 2014