An approach for edge matching large-area satellite image classifications

Michael A. Wulder, Tian Han, Joanne C. White, Christopher R. Butson, and Ronald J. Hall

Abstract. Large-area land cover mapping based on remotely sensed data often requires combining individual or large groups of classified images to produce final map products. Operational and logistical considerations are typically confronted when classifying medium spatial resolution satellite imagery (i.e., Landsat), with the mapping partitioned by spectral, ecological, or political considerations, or combinations thereof. Visual discontinuities can emerge at the locations where logistically based production zones join. Transparent and systematic approaches for addressing discontinuities are desired for the Earth Observation for Sustainable Development of Forests (EOSD) project. This large-area land cover mapping project is producing map products for Canada's forested ecozones. A distributed implementation plan, largely based on grouping provincial and territorial political units, was followed for production. Scene-to-scene discontinuities are rare within each production zone and are primarily related to image acquisition date and phenological state. In contrast, discontinuities can emerge at the production zone boundaries because of differences in support data available or more commonly because of differences in the attribution of density classes. Of the over 475 scenes classified for the EOSD project, it is estimated that fewer than 30 (about 6.3% of total) will require processing to minimize the cross-boundary discontinuity. Options for mitigating the discontinuities are described and demonstrated in the context of different scenarios of overlap found along the EOSD production zone boundaries (complete overlap, partial overlap, and no overlap) using two subsets of a Landsat scene along the shared provincial border between British Columbia and Alberta, Canada. Analysis of image gradients provides a quantitative basis for identification of discontinuities and also relates the results of the likelihood-based relabelling process. Through this process, only density descriptors of cover types are altered, largely maintaining classification integrity. The process as presented is generic and is suitable for addressing edge discontinuities that can emerge when undertaking a large-area land cover classification project.

Résumé. La cartographie du couvert à grande échelle basée sur les données de télédétection requiert souvent la combinaison d'images individuelles ou de groupes importants d'images classifiées pour réaliser des produits cartographiques finaux. Des considérations opérationnelles et logistiques apparaissent généralement lorsque l'on classifie des images satellitaires à résolution spatiale moyenne (c.-à-d., Landsat), la cartographie étant divisée généralement en fonction de motifs logistiques qui sont basés sur des considérations spectrales, écologiques ou politiques ou encore des combinaisons de ces dernières. Aux endroits où les zones de production basées sur la logistique se superposent, des discontinuités visuelles peuvent se manifester. Dans le cadre du projet EOSD (« Earth Observation for Sustainable Development of Forests »), on recherche des approches transparentes et systématiques pour régler ce problème des discontinuités. Ce projet de cartographie à grande échelle du couvert développe des produits cartographiques pour les écozones forestières du Canada. Un plan de mise en place distribué, basé largement sur le regroupement des unités politiques provinciales et territoriales, a été suivi dans le processus de production. À l'intérieur de chacune des zones de production, des discontinuités d'image à image sont plutôt rares et généralement reliées à la date d'acquisition des images et à l'état phénologique associé. Au contraire, au niveau des frontières des zones de production, des discontinuités peuvent survenir à cause des différences dans les données de support disponibles ou, plus souvent, à cause des différences dans l'attribution des classes de densité. Parmi plus de 475 images classifiées, il est prévu que moins de 30 (environ 6,3 % du total) demanderont un traitement pour minimiser la discontinuité au-delà des frontières. On décrit et on démontre les options pour minimiser les discontinuités dans le contexte des différents scénarios de superposition trouvés le long des frontières des zones de production EOSD (superposition totale, superposition partielle et aucune superposition) en utilisant deux sous-

---

M.A. Wulder,[1] T. Han,[2] J.C. White, and C.R. Butson.[3] Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, 506 West Burnside Road, Victoria, BC V8Z 1M5, Canada.

R.J. Hall. Northern Forestry Centre, Canadian Forest Service, Natural Resources Canada, 5320-122 Street, Edmonton, AB T6H 3S5, Canada.

[1]Corresponding author (e-mail: mwulder@nrcan.gc.ca).
[2]Present address: Base Mapping and Data Exchange Unit, Integrated Land Management Bureau, 395 Waterfront Crescent, Victoria, BC V8T 5K7, Canada.
[3]Present address: Forest Analysis and Inventory Branch, British Columbia Ministry of Forests and Range, 7th Floor, 727 Fisgard Street, Victoria, BC V8W 1R8, Canada.

ensembles d'une image de Landsat acquises le long de la frontière commune entre la Colombie-britannique et l'Alberta, au Canada. L'analyse des gradients d'images fournit une base quantitative pour l'identification des discontinuités et permet également de faire le lien avec les résultats de la procédure de ré-étiquetage basée sur la vraisemblance. Par cette procédure, seuls les descripteurs de la densité des types de couvert sont modifiés, assurant globalement l'intégrité de la classification. La procédure, telle que présentée, est générique et adéquate pour régler le problème des discontinuités de contours qui peuvent se manifester lorsque l'on réalise un projet de classification du couvert à grande échelle.
[Traduit par la Rédaction]

Introduction

Large-area land cover mapping based on remotely sensed data sources is a common approach to produce land cover and land-use information to support forest inventory monitoring programmes (Wulder et al., 2003), climate and productivity modelling (Sellers et al., 1996; Liu et al., 1997), and environmental change activities (Townshend et al., 1994). Land cover maps generated from satellite data can be produced over a range of spatial resolutions covering geographical areas of regions (Lunetta et al., 2002; Wessels et al., 2004), nations (Vogelmann et al., 2001; Cihlar et al., 2003), continents (Latifovic et al., 2004), and the globe (DeFries et al., 1998; Hansen et al., 2000). Most national or global land cover products are generated using coarse spatial resolution satellite imagery (>1 km) from sensors that provide data on a daily basis. The large-area coverage of these data reduces the amount of required mosaicking (DeFries and Townshend, 1999; Latifovic et al., 2004). To provide detailed characterization of land surface features, however, mapping land cover at finer spatial resolutions (<1 ha) is in demand by users (Scott and Jennings, 1998) and for reporting purposes (Rosenqvist et al., 2003).

For land cover mapping with medium spatial resolution imagery (e.g., Landsat thematic mapper (TM) or enhanced thematic mapper plus (ETM+)), the smaller pixel size (30 m × 30 m) produces large image files (in terms of physical disk space), with the higher inter-pixel variance. Furthermore, image acquisition priorities, production schedules, and the season or year in which the images were acquired also impact the ease and success with which a large number of Landsat images can be mosaicked together. To address these issues, areas of interest are often partitioned into subareas, defined by ecological or political boundaries (Homer et al., 1997; Scott and Jennings, 1998). Images within these subareas are first combined to create radiometric uniformity and then classified (Homer and Gallant, 2001), or are mosaicked post-classification (Cihlar, 2000).

If multiple images are mosaicked prior to classification, they must be normalized to ensure that the spectral properties of specific land cover classes are consistent across the study area. Recent findings by Olthof et al. (2005) suggest coarse-resolution imagery (Satellite pour l'observation de la terre VEGETATION (SPOT VGT)) can be used as a source of cross-scene radiometric constants for medium spatial resolution imagery (Landsat) that enable the production of radiometrically normalized large-area mosaics, prior to land cover classification. This approach reduced the requirement for edge matching but did not negate the issue completely, as some

boundaries required manual editing post-normalization (Olthof et al., 2005). Other studies have used relative normalization techniques to composite adjacent images without the need for edge matching (Reese et al., 2002) or by merging land cover classes along the seam lines (Han et al., 2004). Guindon (1997) developed a mosaicking approach to accommodate different illumination conditions and the presence of clouds using a grey-level scattergram clustering procedure. By employing a set of reference scenes distributed across the mosaic area, residual errors were minimized when normalizing one scene to the next in a systematic fashion. Vogelmann et al. (2001) mosaicked archived Landsat scenes for the conterminous USA based on leaf-on (summer) and leaf-off (spring) conditions within 31 regional units prior to classification for the National Land Cover Data (NLCD) dataset. In this case, no detail on edge matching was presented, although the authors confirmed that edge matching is a significant issue due to seasonal and interpretation differences along the boundaries of the regional mosaics.

Another approach is to mosaic the classified outputs of each scene (Driese et al., 1997; Fuller et al., 2005). Variations of two separate methods were often used in this regard (Eve and Merchant, 1998). The first involved editing the raster files directly to reconcile discrepancies in land cover labels along adjacent scenes or production zones. The second method involved some form of modelling in the overlap area between two scenes or regions to define a "natural" boundary, based on land cover distributions in the overlap. This type of approach can be labour intensive and largely manual, suggesting it may not be appropriate for projects with large numbers of scenes.

The Earth Observation for Sustainable Development of Forests (EOSD) is an example of a large-area land cover mapping project that requires rectification of class discontinuities at some production zone boundaries. EOSD is a joint project between the Canadian Forest Service (CFS) and the Canadian Space Agency (CSA) with an initial goal of producing a land cover map of Canada representing year 2000 conditions based on Landsat imagery for completion and product delivery in 2006 (Wulder et al., 2003). The imagery available for the project is from a national acquisition and orthorectification program (Wulder et al., 2002) that provided imagery as single scenes towards creation of a national coverage, with the images collected and processed over the period from 1999 to 2005. The EOSD land cover mapping is based on unsupervised *K*-means classification (using the six optical Landsat channels supplemented by a channel capturing the interpixel variance derived from the Landsat panchromatic channel) and subsequent cluster merging and labelling (Wulder

et al., 2004). The EOSD land cover project requires the acquisition, storage, and processing of over 475 Landsat images. The scope of the EOSD program necessitates the involvement and cooperation of many agencies at different jurisdictional levels, institutions, and universities. As a result, consistency and standardization for data processing and classification are critical to the quality of the final map products.

To ensure consistency and standardization, the methodologies, techniques, and procedures involved have been documented and published in reports available for download (http://eosd.cfs.nrcan.gc.ca/cover/index_e.html) and have been followed or adapted by all EOSD participants. A closed and consistent classification legend has also been developed (Wulder and Nelson, 2001) that is compatible in cover type level and class structure to that used by the Canadian National Forest Inventory (NFI) (Gillis et al., 2005). This compatibility facilitates the integration of EOSD, NFI, and other provincial, territorial, and national programs (Remmel et al., 2005).

In an attempt to accommodate limits to project funding and to incorporate ongoing land cover mapping activities being undertaken by provincial and territorial jurisdictions, in some instances (e.g., Alberta, Ontario) the inclusion of provincial mapping projects in EOSD is undertaken. In the case of Alberta, the Alberta Ground Cover Classification (AGCC) (Sánchez-Azofeifa et al., 2005) was ongoing, and this provided the opportunity to translate the AGCC classification to the EOSD class legend. With the Alberta and Ontario land cover mapping projects, the approaches used were complementary, with a crosswalk of the classifications to the EOSD legend (Wulder and Nelson, 2001).

Objective

To leverage local knowledge of land cover characteristics and facilitate linkages with local (provincial or territorial) mapping agencies, the EOSD classification implementation is distributed over four regional CFS centres, including the Pacific Forestry Centre (PFC), Northern Forestry Centre (NoFC), Laurentian Forestry Centre (LFC), and Atlantic Forestry Centre (AFC). Efforts have been made to maintain consistent methodologies across implementation centres. In some cases, however, visual discrepancies in the classification produced along the production zone boundaries between provinces and territories are evident. From review of previous multiscene large-area mapping projects (e.g., Driese et al., 1997; Eve and Merchant, 1998; Vogelmann et al., 2001; Han et al., 2004; Fuller et al., 2005), this issue is expected. The existence of these class discontinuities can degrade the visual effect of the map product and may also introduce artefacts to subsequent applications (Thogmartin et al., 2004). A process is therefore desired for accommodating these class discontinuities.

The EOSD legend (**Table 1**) is developed to be nested in the NFI classification hierarchy, with the selected classes being appropriate for classification with the spatial and spectral properties of Landsat imagery (Wulder and Nelson, 2001). The

level 4 of the NFI relates cover types (e.g., coniferous, deciduous), and level 5 relates increasing detail on density (crown closure) of the cover types. Level 4 classes can be mapped with greater confidence than the level 5 density classes (Remmel et al., 2005; Wulder et al., 2007). Most of the discrepancies found between the classifications of neighbouring production zones are based on different interpretations and assignment of density classes.

The objective of this communication is to recommend and demonstrate a quantitative approach for identification of a discontinuity, and to provide a solution for mitigation, including the application to differing overlap scenarios (full, partial, and none). The reassignment of the density class labels for a given cover type in a transparent and automated fashion is developed and demonstrated. The workflow and subsequent rectification of edge class discontinuities are generic and are applicable beyond the program and examples presented in this paper.

Study area

The scope of the EOSD project is presented in **Figure 1**, including the Landsat frames required to be classified (>475) and the production zones covered by different CFS centres. The EOSD land cover mapping project is focused on Canada's forested ecozones, covering over 60% of Canada's landmass. When including image extension outside of the forested ecozones, over 80% (or 8 million square kilometres) of Canada is mapped (Wulder et al., 2006). A focus area, the frame (path 48, row 20, hereafter noted as 48/20) outlined in cyan in **Figure 1**, has been selected to demonstrate the method applied and subsequent outcome. The focus area is on the jurisdictional boundary between British Columbia (BC) and Alberta (AB), Canada, and is within the Boreal Plain ecozone, which is typified by flat topography and a mix of coniferous and broadleaf forests and wetland (Ecological Stratification Working Group, 1995).

For large-area land cover mapping based on medium spatial resolution images, map products are often produced by mosaicking either input images before classification or the classified results of individual input images. Taking the latter approach, the EOSD team classified individual Landsat images and then combined them to generate land cover products.

Within each of the EOSD production zones (**Figure 1**), classification discontinuity between adjacent scenes is negligible, largely attributable to the two-stage approach employed in the EOSD data processing (Wulder et al., 2004). The first stage of the approach is the preclassification image processing (PIP), including the radiometric correction of source imagery to top-of-atmosphere (TOA) reflectance (Chander and Markham, 2003). The second stage is the postclassification edge matching (PEM), which is characterized using the results of a completed classification of one scene to inform the cluster labelling for subsequent scenes.

Landsat ETM+ scene 48/20 was acquired on 4 August 1999 and is located on the boundary between the PFC and NoFC

Table 1. EOSD class legend.

CODE	COLOR	CLASS	DESCRIPTION
0		No Data	
12		Shadow	
11		Cloud	
31		Snow/Ice	Glacier/snow
32		Rock/Rubble	Bedrock, rubble, talus, blockfield, rubbly mine spoils, or lava beds
33		Exposed Land	River sediments, exposed soils, pond or lake sediments, reservoir margins, beaches, landings, burned areas, road surfaces, mudflat sediments, cutbanks, moraines, gravel pits, tailings, railway surfaces, buildings and parking, or other non-vegetated surfaces
20		Water	Lakes, reservoirs, rivers, streams, or salt water
51		Shrub - Tall	At least 20% ground cover which is at least one-third shrub; average shrub height greater than or equal to 2 m
52		Shrub - Low	At least 20% ground cover which is at least one-third shrub; average shrub height less than 2 m
100		Herb	Vascular plant without woody stem (grasses, crops, forbs, gramminoids); minimum of 20% ground cover or one-third of total vegetation must be herb
40		Bryoids	Bryophytes (mosses, liverworts, and hornworts) and lichen (foliose or fruticose; not crustose); minimum of 20% ground cover or one-third of total vegetation must be a bryophyte or lichen
81		Wetland - Treed	Land with a water table near/at/above soil surface for enough time to promote wetland or aquatic processes; the majority of vegetation is coniferous, broadleaf, or mixed wood
82		Wetland - Shrub	Land with a water table near/at/above soil surface for enough time to promote wetland or aquatic processes; the majority of vegetation is tall, low, or a mixture of tall and low shrub
83		Wetland - Herb	Land with a water table near/at/above soil surface for enough time to promote wetland or aquatic processes; the majority of vegetation is herb
211		Coniferous - Dense	Greater than 60% crown closure; coniferous trees are 75% or more of total basal area
212		Coniferous - Open	26-60% crown closure; coniferous trees are 75% or more of total basal area
213		Coniferous - Sparse	10-25% crown closure; coniferous trees are 75% or more of total basal area
221		Broadleaf - Dense	Greater than 60% crown closure; broadleaf trees are 75% or more of total basal area
222		Broadleaf - Open	26-60% crown closure; broadleaf trees are 75% or more of total basal area
223		Broadleaf - Sparse	10-25% crown closure; broadleaf trees are 75% or more of total basal area
231		Mixed Wood - Dense	Greater than 60% crown closure; neither coniferous nor broadleaf trees account for 75% or more of total basal area
232		Mixed Wood - Open	26-60% crown closure; neither coniferous nor broadleaf trees account for 75% or more of total basal area
233		Mixed Wood - Sparse	10-25% crown closure; neither coniferous nor broadleaf trees account for 75% or more of total basal area

production zones. The image was classified using the class legend shown in **Table 1** and EOSD methods on the BC side. On the AB side, the image was classified using similar procedures that are part of the AGCC process (Sánchez-Azofeifa et al., 2005). When the classifications are directly mosaicked using the jurisdictional boundary between BC and AB as the mosaic cut-line, classification discontinuity is marked as shown in panels A and B in **Figure 2**, where the classifications on the left in both panels are done by PFC and those on the right were undertaken by the University of Alberta as part of the AGCC and then translated to the EOSD class structure (hereinafter referred to as NoFC classification).

Figure 1. EOSD land cover production zones and study area (noted with the cyan outline).
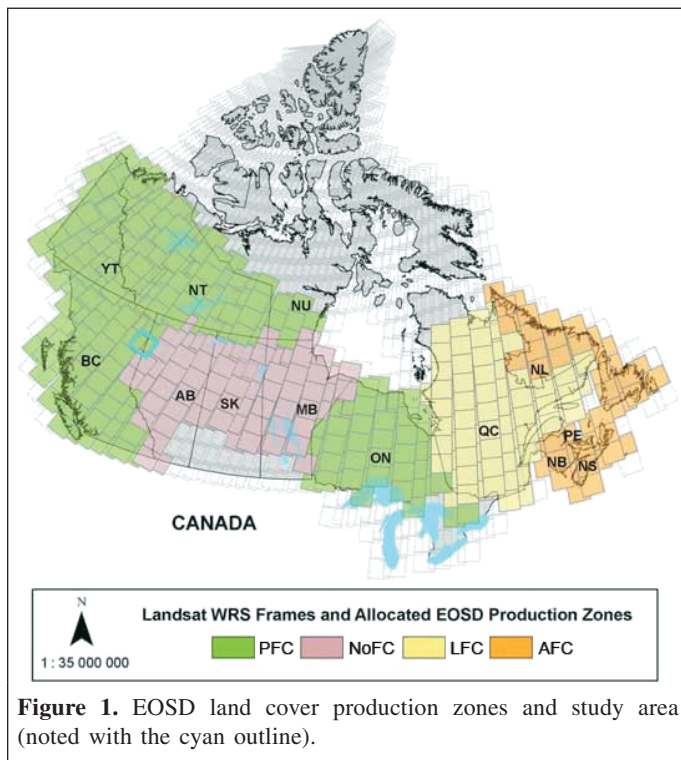
Methodology

The rationale of the proposed approach is to reduce the classification discontinuity by relabelling the pixels responsible for the discontinuity from one class to another, with the steps outlined in **Figure 3**. To minimize the impact on overall classification, the relabelling is restricted to areas adjacent to the boundaries between production zones where the discontinuity is apparent and to complementary categories (such as the density classes for a given cover type). Three inputs are required to initiate the approach, including two classified images of adjacent production zones and the boundary line that separates the two production zones. The main steps include identifying the classes causing the discontinuity, determining the control and dependent classified images, relabelling the pixels, and quantitatively measuring the discontinuity before and after the discontinuity is treated. These steps are detailed in the following subsections.

Identification of classes causing discontinuity

Prior to edge matching, an investigation as to the nature of the classification discontinuity is required. The investigation is conducted in the overlap of the two input images where histograms are calculated and compared. This enables identification of which classes appear to be mapped preferentially different on either side of the discontinuity. The overlap of the input images may also be used as a guide for the area eligible for pixel relabelling. If no overlap is available, a buffer across the production zone boundary may be constructed

to serve as the region of interest where histograms are calculated.

Typically, three overlap scenarios exist for the input classified images in the EOSD project, namely complete, partial, and none. As only part of the overlap is required by the proposed edge-matching approach, complete overlap will be treated the same as partial overlap, which reduces the number of scenarios from three to two. The available classified images for 48/20 are partially overlapped. To enable demonstration of both overlap and non-overlap cases, two spatial subsets are selected from scene 48/20 as indicated in **Figure 2**, where the subset indicated by panel A is chosen to represent the partial overlap scenario and the subset indicated by panel B is used to represent the non-overlap scenario, of which the overlap was deliberately removed.

To identify the classes causing the discontinuity, a histogram was created for both NoFC and PFC classified images within a specified area. For the partial overlap case indicated by panel A in **Figure 2**, the specified area is selected as the overlap on the left of the boundary (within the BC side) where a histogram is created for each input classified image (**Figure 4a**). For the non-overlap case shown in panel B of **Figure 2**, a buffer is created along the boundary between BC and AB which is used as the specified area. The width of the buffer on each side of the boundary is set to ~8 km. The histogram of the NoFC classified image is calculated using the buffered pixels on the AB side, which is compared to the histogram of the PFC classified image produced using the buffered pixels on the BC side (**Figure 4b**).

Comparing these histograms indicates that the coniferous dense (class 211) and open (class 212) are the two classes that differ most in both cases. According to the NoFC classification, there are ~95 000 coniferous dense and 0 coniferous open pixels for the overlap case (panel A of **Figure 2**) and ~70 000 coniferous dense and 0 coniferous open pixels for the non-overlap case (panel B of **Figure 2**). According to the PFC classification, however, the corresponding numbers are ~4000 and ~100 000 for the overlap case and ~1000 and 75 000 for the non-overlap case. It is anticipated that the dramatic class discrepancy caused the visual discontinuity when the classified images by NoFC and PFC are mosaicked along the boundary. The crosswalk of the EOSD and AGCC legends (Wulder and Nelson 2001) reveals a difference in the number and nature of density classes. The AGCC classifications depict two density levels, whereas the EOSD mapping attempts to capture three density levels.

In addition to the class pair of coniferous dense and open, which is considered as the primary contributor to the classification discontinuity, there are other class pairs, including broadleaf dense (class 221) and open (class 222) and mixed wood dense (class 231) and open (class 232), which contribute secondarily to the discontinuity. It is noted that the two classes within each pair are complementary, i.e., the same cover type but different density.
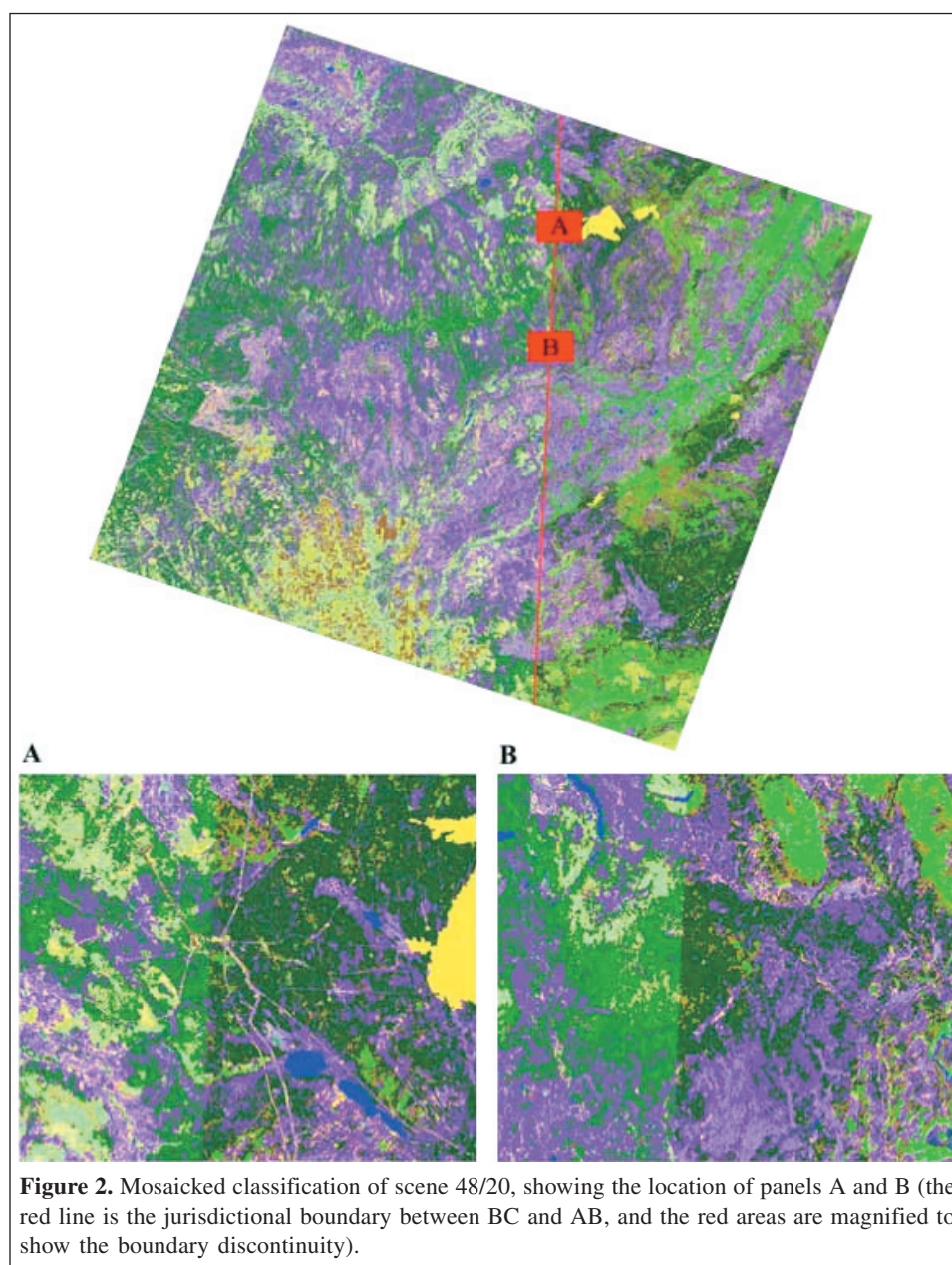
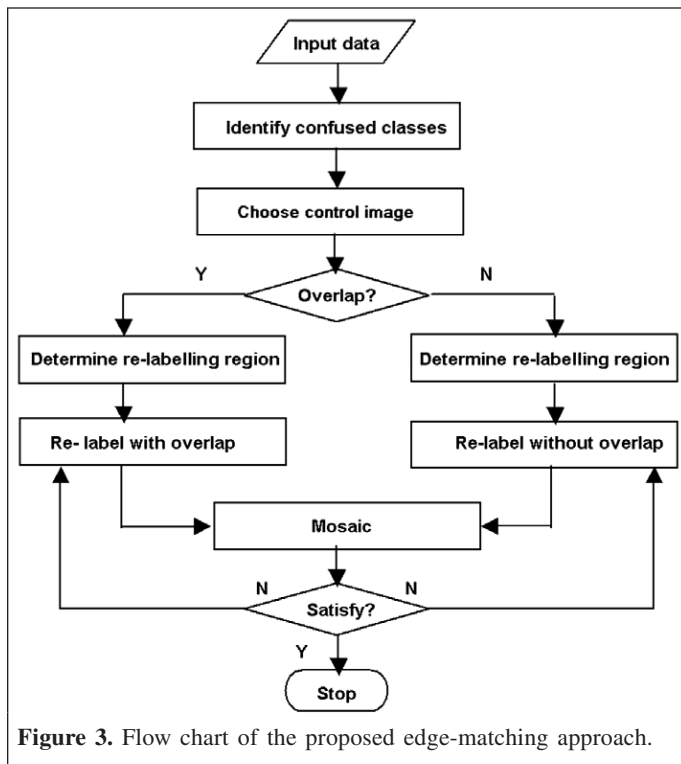### Determination of control and dependent images

In the proposed approach, the two input classified images produced by NoFC and PFC play different roles, whereby the characteristics of one classification are considered preferable, with one considered the control and the other dependent. The determination of control or dependent designation is based on several factors, including the availability of training data, the image date, and the presence of haze or cloud. Based on greater access to project-specific ground data, the NoFC classified image is selected in this case as the control, leaving the PFC image as the dependent. For the overlap scenario indicated by panel A of **Figure 2**, the overlapped pixels of certain classes in the dependent classified image are selectively relabelled to match the classes in the control classified image. Following the same strategy, for the non-overlap scenario indicated by panel

B of **Figure 2**, the buffered pixels in the dependent are selectively re-relabelled to match the classes in the control.

Pixel relabelling

The proposed approach focuses on relabelling the classes responsible for the discontinuity along the boundary between two production zones. To minimize the impact of the relabelling on the overall classification, the following three restrictions are exercised. The first restriction is on the area subject to relabelling, which is confined to the area of the overlap (overlap case) or the buffer (non-overlap case), both within the dependent image. The second restriction is on the classes selected for relabelling. Since the classification discontinuity is due to the disagreement between density classes, only pixels of these classes are considered to be

Figure 2. Mosaicked classification of scene 48/20, showing the location of panels A and B (the red line is the jurisdictional boundary between BC and AB, and the red areas are magnified to show the boundary discontinuity).

Figure 3. Flow chart of the proposed edge-matching approach.

Figure 4. Histogram for (a) the overlap case, and (b) the non-overlap case.
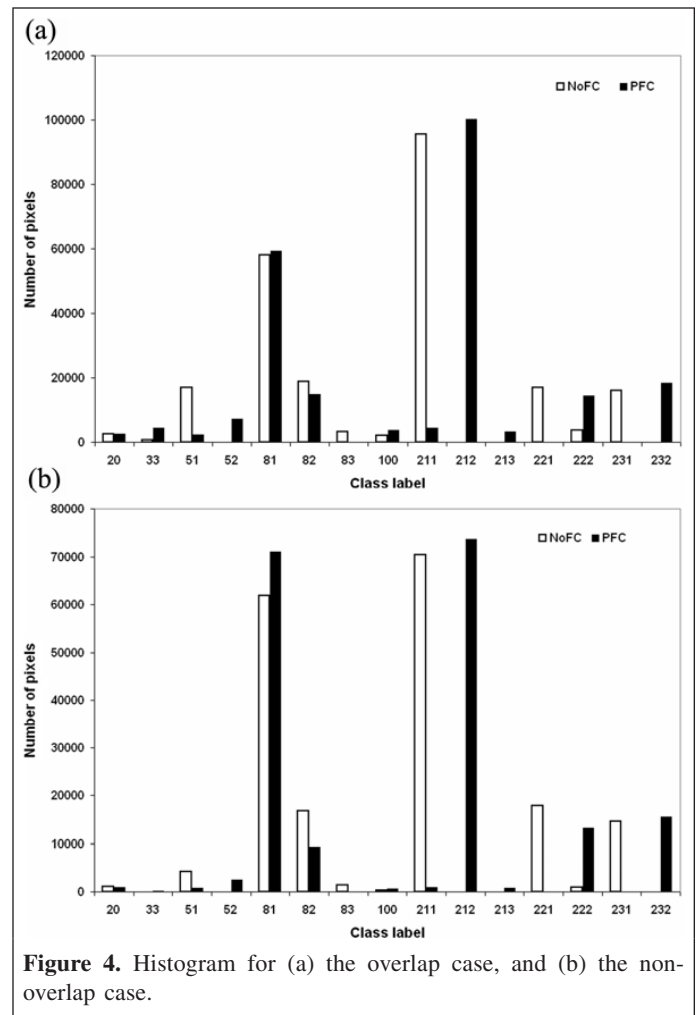
relabelled (from one density class to another of the same type). The final restriction is on the distance between a pixel position and the boundary. The closer a pixel is located to the boundary, the greater the likelihood that the pixel will be relabelled.

The idea of this pixel relabelling is illustrated row-wise in **Figure 5**, where the color-coded squares represent pixels of different class. As **Figure 5** shows, the control classification assigned most of the pixels to blue, and the dependent classification assigned most of the pixels to green, which results in discontinuity when the mosaicking was conducted along the jurisdictional boundary. To reduce the discontinuity, the green pixels in the specified area between the two red bars (overlap (buffer) and jurisdictional boundaries) are selectively relabelled to blue. Transition zones within the specified area are created, of which each contains an approximately equal number of pixels. For the transition zone that is far from the jurisdictional boundary, such as zone n in **Figure 5**, its green pixels are given a small likelihood of being relabelled to blue. As the transition zone moves closer to the jurisdictional boundary, its green pixels are given more chance of being relabelled. The relabelling is carried out top-down on a row basis.

Quantitative measure for discontinuity

The relabelling needs to be conducted iteratively, as each round of relabelling targets a particular class in question (e.g., coniferous open). After each relabelling, the relabelled dependent and the unchanged control classified image are mosaicked. The output is inspected to determine if another round of relabelling is required. Questions may arise at this point: What is the stopping rule for the iteration and is there any quantitative measure to evaluate the effectiveness of the proposed edge-matching approach? Here an image-gradient-based method is devised to address these questions.

Image gradient represents the rate of change of pixel intensities over a local neighbourhood, which is often employed to identify discontinuities and edges in digital image processing (Pitas, 2000). Originally, the gradient of a continuous function $f(x, y)$ is calculated as follows:

$$|\nabla f(x, y)| = \sqrt{\left[\frac{\partial f(x, y)}{\partial x}\right]^2 + \left[\frac{\partial f(x, y)}{\partial y}\right]^2} \tag{1}$$

For digital images, the gradient calculation is simplified as

$$|\nabla f(x, y)| = \sqrt{[f(x+1, y) - f(x, y)]^2 + [f(x, y+1) - f(x, y)]^2} \tag{2}$$

where x and y indicate the location (row, column) of the pixel being considered. To accommodate the EOSD classified images, the following modifications were made for the gradient calculation. First, only the horizontal gradient (x, $x + 1$ direction) is calculated (based on the nature of the jurisdictional
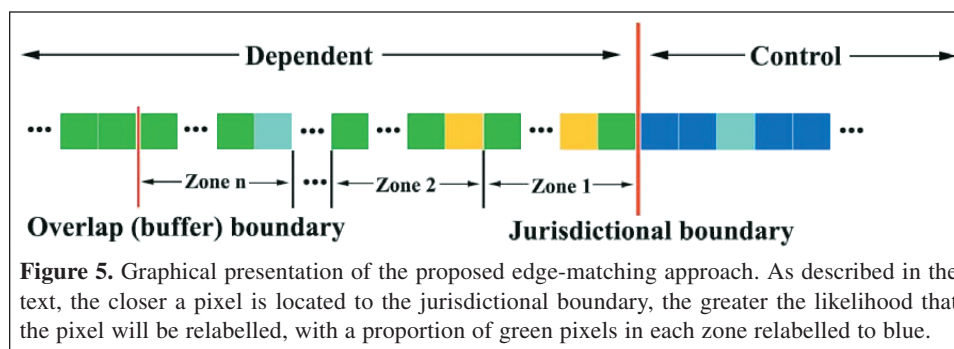
Figure 5. Graphical presentation of the proposed edge-matching approach. As described in the text, the closer a pixel is located to the jurisdictional boundary, the greater the likelihood that the pixel will be relabelled, with a proportion of green pixels in each zone relabelled to blue.

boundary in this example). Second, to mitigate the effect of class labels on gradient calculation, all class labels are treated equally, i.e., the gradient difference of two pixels of the same class is 0, otherwise 1. Lastly, the gradient is calculated by following the boundary. The absolute gradient difference between a pixel and the one immediately to the left is computed and totalled up across all rows of the input classified image. These modifications result in the following formula that is used to calculate gradient:

$$g_j = \sum_{i=1}^{n} |f(i, j+1) - f(i, j)| \tag{3}$$

where g_j is the gradient calculated at the position j pixels way from the boundary, and therefore g_0 is the gradient calculated on the boundary; and n is the number of rows of the mosaicked classified image.

It is anticipated that the gradient calculated this way peaks on the boundary, i.e., g_0 is the maximum across all columns of the mosaicked classified image. This can be used as a quantitative indicator to evaluate the effectiveness of the proposed edge-matching approach and to establish the stopping rule for the iteration of pixel relabelling. For example, if the gradient calculated on the boundary (g_0) is no longer the maximum across the image, then the iteration of relabelling may be terminated.

Results and discussion

As indicated in the previous sections, the classes that contribute to the classification discontinuity along the production zone boundaries are a result of differing trends in the attribution of vegetation density classes, specifically coniferous dense (class 211) and open (class 212). The cover types are not altered, only the attribution of density class. To quantify the discontinuity, gradient values are calculated using Equation (3) across the boundary with 100 pixels on each side of the boundary. As expected, for both overlap and non-overlap cases represented by panels A and B, respectively, in **Figure 2** the gradient values calculated on the boundary are markedly greater than those calculated elsewhere, indicated by the pre-relabelling gradient curves in **Figures 6** and **7**.

To resolve the identified discontinuity, the proposed edge-matching approach is implemented with three transition zones. Starting with the overlap case where the boundary discontinuity is primarily due to the inconsistent classification of coniferous dense (class 211) and open (class 212) between NoFC and PFC as indicated in **Figure 4a**, the proposed approach uses the control class label (coniferous dense) to relabel the dependent pixels of coniferous open in the overlap area on the BC side. The relabelling is effective gradient-wise (first relabelling in **Figure 6**, where gradient is reduced from 0.84 to 0.61) and visually (**Figure 8a**, where the relabelled images are mosaicked). It is apparent that the discontinuity has been greatly reduced. A closer look, however, reveals that there is a marked broken point of gradient on the boundary, and the discontinuity is still faintly visible in the mosaicked image. To further reduce the discontinuity, the second relabelling is conducted to relabel the mixed wood open (class 232) as dense (class 231). As shown in **Figure 4a**, the inconsistent classification of these two classes is the secondary contributor to the discontinuity. After this relabelling, the gradient on the boundary is further reduced to 0.47 (second relabelling in **Figure 6**) and the discontinuity is nearly invisible (**Figure 8b**). The gradient curve of the second relabelling also shows that a smooth transition has been established across the boundary. As the gradient on the boundary is no longer the maximum, this relabelling process may stop after the second relabelling.

The non-overlap case represented by panel B of **Figure 2** is processed by following a similar approach. The only difference is that the classes involved in the second relabelling are broadleaf dense (class 221) and open (class 222). Pixels of broadleaf open are relabelled as broadleaf dense. The gradients of pre- and post-relabelling are shown in **Figure 7**, of which the gradients calculated on the boundary are 0.80, 0.53, and 0.37 for the pre-relabelling, and for the first and second relabelling, respectively. The improvement of the relabelling is also demonstrated in **Figure 9**, where **Figures 9a** and **9b** are the mosaics of the input images after the first and second relabellings, respectively.

The proposed edge-matching approach has several advantages. For example, it only requires as input the classified images and the boundary line between them. Preventive measures are employed to minimize the impact of pixel relabelling on the entire classified image. The relabelling is confined to the overlap or buffered area on the dependent image
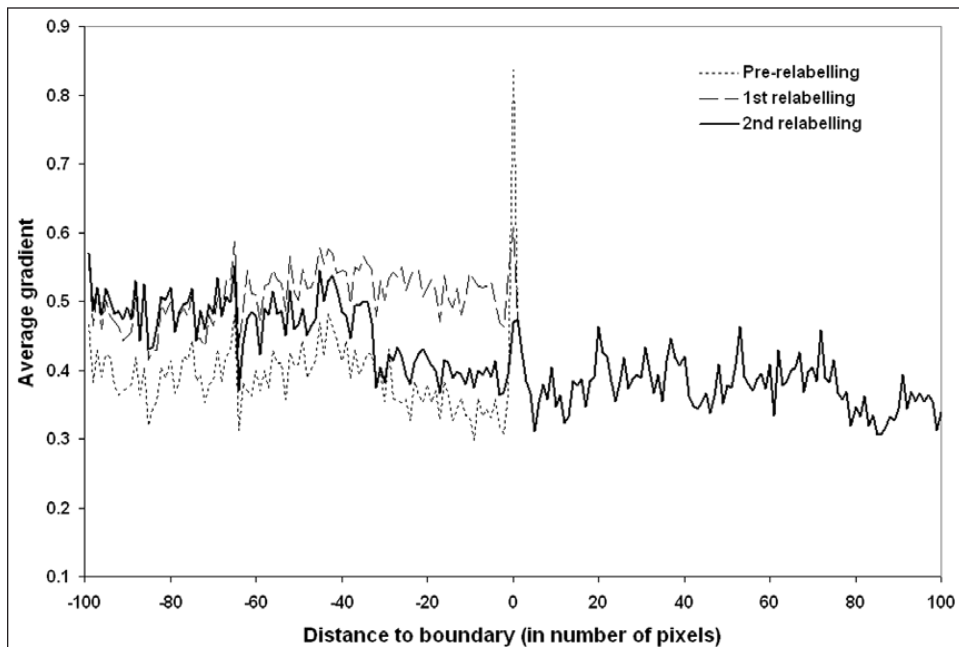
Figure 6. Gradients calculated across boundary for the overlap case (the negative numbers on the *x* axis indicate the positions west of the boundary, and the positive numbers indicate those east of the boundary).

Figure 7. Gradients calculated across boundary for the non-overlap case (the negative numbers on the *x* axis indicate the positions west of the boundary, and the positive numbers indicate those east of the boundary).
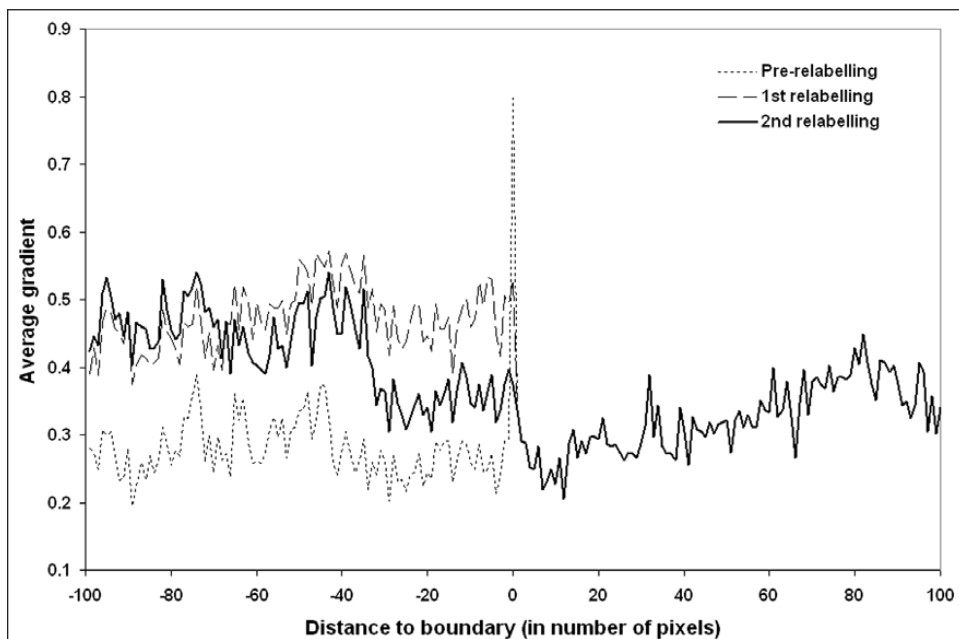
side. Pixels outside this area are not altered in any way. In addition, only pixels of the specifically identified classes are selectively relabelled. The approach can be implemented in a semi-automatic fashion, where the main operator intervention is to decide which classes are to be relabelled. The success of pixel relabelling can be evaluated not only visually but also quantitatively in terms of gradient. All data processing is raster-based and can be implemented to handle large images.

The proposed edge-matching method has been implemented using interactive data language (IDL). Running on a SUN4U workstation with 4 gigabytes of memory, the IDL code only took about 40 s to process the entire Landsat scene 48/20 using
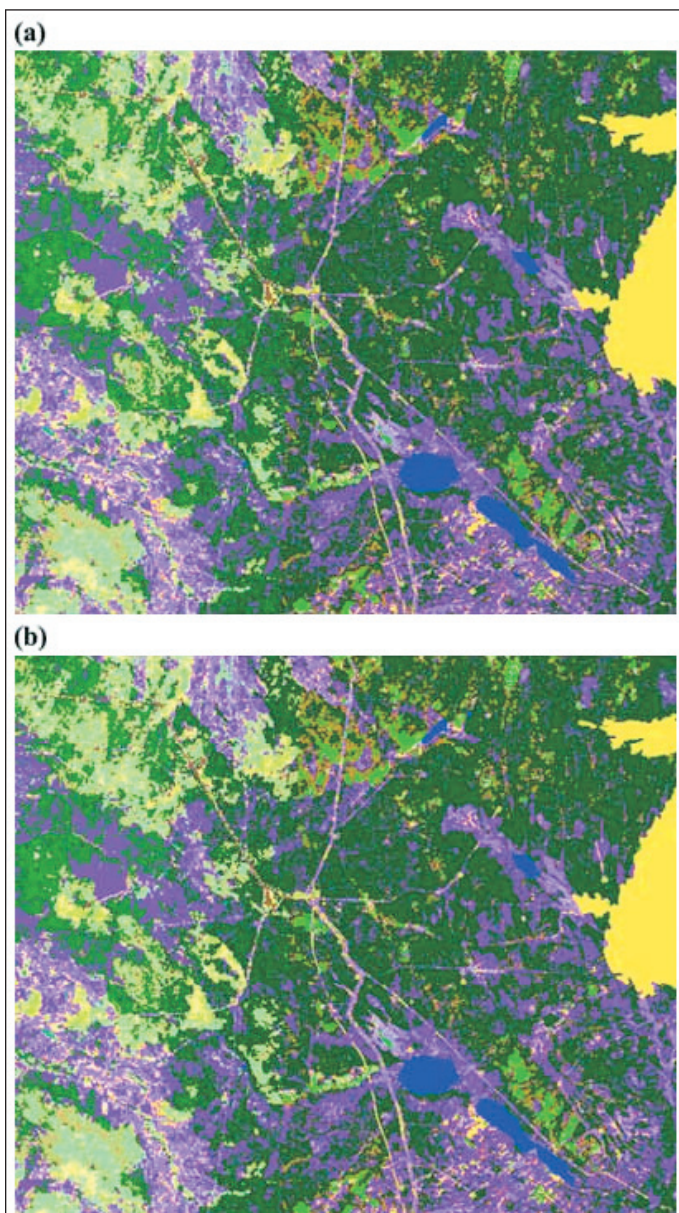
Figure 8. Mosaic after (a) first relabelling, and (b) second relabelling (overlap case).
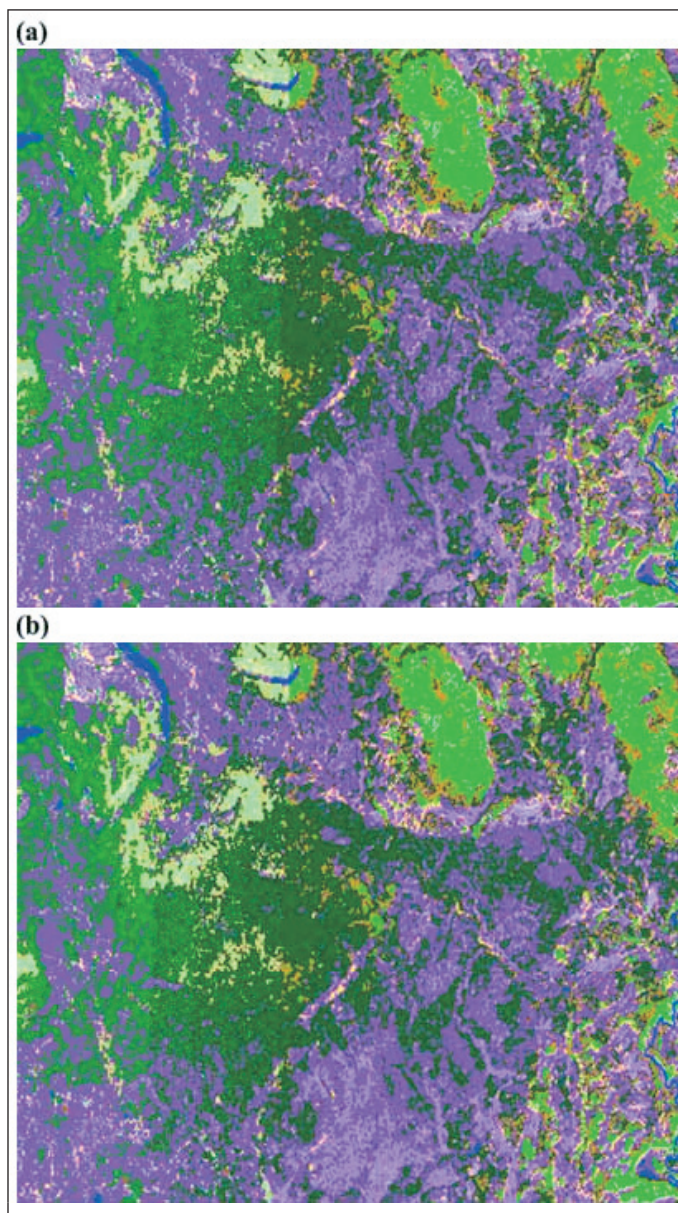

Figure 9. Mosaic after (a) first relabelling, and (b) second relabelling (non-overlap case).

two rounds of relabelling. The mosaic of the processed input classified images is shown in **Figure 10**. Similar results have been found for other boundary locations as exemplified earlier. Further, the algorithm can also be applied on boundaries with an orientation other than the north–south situation presented here.

The edge-matching approach may also be used as a means for screening classified images with an intention to revisit and relabel, to ensure the class discontinuities present are not due to classification error. Although such an approach is not appropriate for a mapping exercise as large as EOSD, province- or state-wide projects may have the funds and support data to enable such an option. The difficulties in mapping forest density (crown closure) have been illustrated by which classes

were most frequently subject to relabelling. Options to determine crown closure via alternate approaches, including regression-based approaches or continuous field mapping, for projects that have access to the required support data could also be considered.

Conclusions

Edge matching is a useful option for large-area land cover mapping projects that use medium spatial resolution satellite imagery. Edge-class anomalies have implications for the end users of the data and subsequent applications. Approaches for alleviating classification discontinuities along production zone boundaries can be implemented before and (or) after image
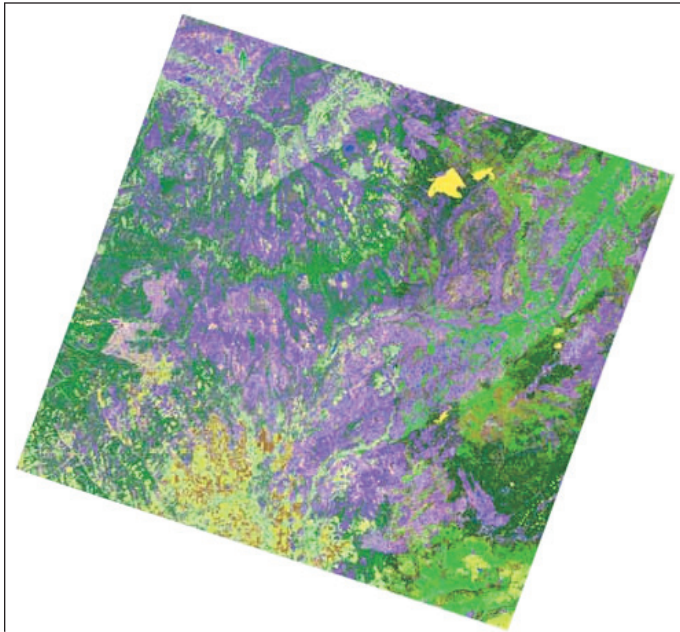
Figure 10. Mosaic of scene 48/20 after second relabelling.

classification. Images may be normalized and mosaicked, and (or) production zones may be used to partition the study area into homogenous subunits prior to classification. Well-documented methods for image processing and classification, combined with good communication and coordination between classification teams, also help to reduce the occurrence of the discontinuities. Although large-area satellite-based mapping projects are increasingly undertaken, the approaches followed for edge matching are not well documented in the literature. Transparent, systematic, and repeatable approaches for addressing edge-matching needs of large-area land cover mapping projects are accordingly desired.

An approach for edge matching large-area image classifications has been presented in this paper. The framework involves identification of the nature of the edge (discontinuity) across the boundary between the two production zones, determination of a control and dependent image, and the implementation of a relabelling protocol. The process is iterative and has restrictions designed to minimize the impact on the overall classification integrity. A quantitative measure is proposed based on the revised image gradient to supplement the use of visual inspection for identifying and resolving boundary discontinuity.

This study used the EOSD classified images of two subset areas within a Landsat scene between British Columbia and Alberta to demonstrate the effectiveness of the approach. The results produced showed that the proposed edge-matching approach can reduce the classification discontinuity along production zone boundaries. Final EOSD production stages are implementing this approach to process the scenes where required. This approach is generic, repeatable, and flexible and may be implemented semi-automatically for large-area land cover mapping projects.

Acknowledgements

References

Chander, G., and Markham, B. 2003. Revised Landsat-5 TM radiometric calibration procedures and post-calibration dynamic ranges. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, pp. 2674–2677.

Cihlar, J. 2000. Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing*, Vol. 21, pp. 1093–1114.

Cihlar, J., Latifovic, R., Beaubien, J., Guindon, B., and Palmer, M. 2003. Thematic mapper (TM) based accuracy assessment of a land cover product for Canada derived from SPOT VEGETATION (VGT) data. *Canadian Journal of Remote Sensing*, Vol. 29, No. 2, pp. 154–170.

DeFries, R.S., and Townshend, J.R.G. 1999. Global land cover characterization from satellite data: From research to operational implementation? *Global Ecology and Biogeography*, Vol. 8, pp. 367–379.

DeFries, R.S., Hansen, M., Townshend, J.R.G., and Sohlberg, R. 1998. Global land-cover classification at 8 km spatial resolution: Part 1. Training and validation data derived from Landsat imagery. *International Journal of Remote Sensing*, Vol. 19, pp. 3141–3168.

Driese, K.L., Reiners, W.A., Merrill, E.H., and Gerow, K.G. 1997. A digital land cover map of Wyoming, USA: a tool for vegetation analysis. *Journal of Vegetation Science*, Vol. 8, pp. 133–146.

Ecological Stratification Working Group. 1995. A *national ecological framework for Canada*. Centre for Land and Biological Resources Research, Research Branch, Agriculture and Agri-Food Canada, Ottawa, and Ecozone Analysis Branch, State of the Environment Directorate, Environment Canada, Ottawa. Report and national map, scale 1 : 7 500 000.

Eve, M.D., and Merchant, J.W. 1998. *A national survey of land cover mapping protocols used in the GAP analysis program: Final Report*. Available from http://calmit.unl.edu/gapmap/ [accessed 15 January 2007].

Fuller, R.M., Cox, R., Clarke, R.T., Rothery, P., Hill, R.A., Smith, G.M., Thomson, A.G., Brown, N.J., Howard, D.C., and Stott, A.P. 2005. The UK land cover map 2000: Planning, construction, and calibration of a remotely sensed, user-oriented map of broad habitats. *International Journal of Applied Earth Observation and Geoinformation*, Vol. 7, pp. 202–216.

Gillis, M.D., Omule, A.Y., and Brierley, T. 2005. Monitoring Canada's forests: the National Forest Inventory. *The Forestry Chronicle*, Vol. 81, pp. 214–221.

Guindon, B. 1997. Assessing the radiometric fidelity of high resolution satellite image mosaics. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 52, pp. 229–243.

Han, K-S., Champeaux, J-L., and Roujean, J-L. 2004. A land cover classification product over France at 1 km resolution using SPOT4/VEGETATION data. *Remote Sensing of Environment*, Vol. 91, pp. 52–66.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., and Sohlberg, R. 2000. Global land-cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, Vol. 21, pp. 1331–1364.

Homer, C.G., and Gallant, A. 2001. *Partitioning the conterminous United States into mapping zones for Landsat TM land cover mapping*. US Geological Survey, Draft White Paper. Available from http://landcover.usgs.gov/publications.asp. [accessed 29 March 2007].

Homer, C.G., Ramsey, R.D., Edwards, T.C., Jr., and Falconer, A. 1997. Landscape cover-type modeling using a multi-scene thematic mapper mosaic. *Photogrammetric Engineering & Remote Sensing*, Vol. 63, pp. 59–67.

Latifovic, R., Zhu, Z., Cihlar, J., Giri, C., and Olthof, I. 2004. Land-cover mapping of north and central America — global land cover 2000. *Remote Sensing of Environment*, Vol. 89, pp. 116–127.

Liu, J., Chen, J.M., Cihlar, J., and Park., W. 1997. A process-based boreal ecosystem productivity simulator using remote sensing inputs. *Remote Sensing of Environment*, Vol. 62, pp. 158–175.

Lunetta, R.S., Ediriwickrema, J., Johnson, D., Lyon, J.G., and McKerrow, A. 2002. Impacts of vegetation dynamics on the identification of land-cover change in a biologically complex community in North Carolina, USA. *Remote Sensing of Environment*, Vol. 82, pp. 258–270.

Olthof, I., Butson, C., and Fraser, R. 2005. Signature extension through space for northern landcover classification: A comparison of radiometric correction methods. *Remote Sensing of Environment*, Vol. 95, pp. 290–302.

Pitas, I. 2000. *Digital image processing algorithms and applications*. John Wiley & Sons, Inc., New York. 418 pp.

Reese, H.M., Lillesand, T.M., Nagel, D.E., Stewart, J.S., Goldmann, R.A., Simmons, T.E., Chipman, J.W., and Tessar, P.A. 2002. Statewide land cover derived from multi-seasonal Landsat TM data: a retrospective of the WISCLAND project. *Remote Sensing of Environment*, Vol. 82, pp. 224–237.

Remmel, T., Csillag, F., Mitchell, S., and Wulder, M.A. 2005. Integration of forest inventory and satellite imagery: a Canadian status assessment and research issues. *Forest Ecology and Management*, Vol. 207, pp. 405–428.

Rosenqvist, A., Milne, A., Lucas, R., Imhoff, M., and Dobson, C. 2003. A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science and Policy*, Vol. 6, pp. 441–455.

Sánchez-Azofeifa, G.A., Chong, M., Sinkwich, J., and Mamet, S. 2005. *Alberta ground cover characterization (AGCC) training and procedures manual*. Earth Observations System Laboratory, Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alta.

Scott, J.M., and Jennings, M.D. 1998. Large-area mapping of biodiversity. *Annals of the Missouri Botanical Garden*, Vol. 85, pp. 34–47.

Sellers, P.J., Randall, D.A., Collatz, G.J., Berry, J.A., Field, C.B., Dazlich, D.A., Zhang, C., Collelo, G.D., and Bounoua, L. 1996. A revised land surface parameterization (SiB2) for atmosphere GCMs — Part I-model formulation. *Journal of Climate*, Vol. 9, pp. 676–705.

Thogmartin, W.E., Gallant, A.L., Knutson, M.G., Fox, T.J., and Suarez, M.J. 2004. A cautionary tale regarding use of the National Landcover Dataset 1992. *Wildlife Society Bulletin*, Vol. 32, pp. 970–978.

Townshend, J.R.G., Justice, C.O., Skole, D., Malingreau, J.-P., Cihlar, J., Teillet, P., Sadowski, F., and Ruttenberg, S. 1994. The 1 km resolution global data set: needs of the International Geosphere–Biosphere Programme. *International Journal of Remote Sensing*, Vol. 15, pp. 3417–3441.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K., and Van Driel, N. 2001. Completion of the 1990s national land cover data set for the conterminous United States from Landsat thematic mapper data and ancillary data sources. *Photogrammetric Engineering & Remote Sensing*, Vol. 67, pp. 650–662.

Wessels., K.J., De Fries, R.S., Dempewolf, J., Anderson, L.O., Hansen, A.J., Powell, S.L., and Moran, E.F. 2004. Mapping regional land cover with MODIS data for biological conservation: examples from the Great Yellowstone Ecosystem, USA and Pará State, Brazil. *Remote Sensing of Environment*, Vol. 92, No. 1, pp. 67–83.

Wulder, M.A., and Nelson, T. 2001. *EOSD legend: characteristics, suitability, and compatibility, version 1*. Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, Victoria, B.C.

Wulder, M., Loubier, E., and Richardson, D. 2002. Landsat-7 ETM+ orthoimage coverage of Canada. *Canadian Journal of Remote Sensing*, Vol. 28, No. 5, pp. 667–671.

Wulder, M.A., Dechka, J.A., Gillis, M.A., Luther, J.E., Hall, R.J., Beaudoin, A., and Franklin, S.E. 2003. Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *The Forestry Chronicle*, Vol. 79, pp. 1075–1083.

Wulder, M.A., Cranny, M., Dechka, J., and White, J.C. 2004. *An illustrated methodology for land cover mapping of forests with Landsat-7 ETM+ data: methods in support of EOSD land cover, version 3*. Pacific Forestry Centre, Canadian Forest Service, Natural Resources Canada, Victoria, B.C. 35 pp.

Wulder, M., Cranny, M., Hall, R., Luther, J., Beaudoin, A., and Dechka, J. 2006. *Satellite land cover mapping of Canada's forests*. SPIE Newsroom, The International Society for Optical Engineering, Bellingham, Wash. 2 pp.

Wulder, M.A., White, J.C., Magnussen, S., and McDonald, S. 2007. Validation of a large area land cover product using purpose-acquired airborne video. *Remote Sensing of Environment*, Vol. 106, No. 4, pp. 480–491.