Representative Landscapes in the Forested Area of Canada

Jeffrey A. Cardille · Joanne C. White ·
Mike A. Wulder · Tara Holland

Abstract Canada is a large nation with forested ecosystems that occupy over 60% of the national land base, and knowledge of the patterns of Canada's land cover is important to proper environmental management of this vast resource. To this end, a circa 2000 Landsat-derived land cover map of the forested ecosystems of Canada has created a new window into understanding the composition and configuration of land cover patterns in forested Canada. Strategies for summarizing such large expanses of land cover are increasingly important, as land managers work to study and preserve distinctive areas, as well as to identify representative examples of current land-cover and land-use assemblages. Meanwhile, the development of extremely efficient clustering algorithms has become increasingly important in the world of computer science, in which billions of pieces of information on the internet are continually sifted for meaning for a vast variety of applications. One recently developed clustering algorithm quickly groups large numbers of items of any type in a given data set while simultaneously selecting a representative—or "exemplar"—from each cluster. In this context, the availability of both advanced data processing methods and a nationally available set of landscape metrics presents an opportunity to identify sets of representative landscapes to better understand landscape pattern, variation, and distribution across the forested area of Canada. In this research, we first identify and provide context for a small, interpretable set of exemplar landscapes that objectively represent land cover in each of Canada's ten forested ecozones. Then, we demonstrate how this approach can be used to identify flagship and satellite long-term study areas inside and outside protected areas in the province of Ontario. These applications aid our understanding of Canada's forest while augmenting its management toolbox, and may signal a broad range of applications for this versatile approach.

Keywords Canada · Forest · EOSD · Fragmentation · Pattern · Exemplars · Representatives · Affinity propagation

Introduction

In recent decades, the merging of GIS and remote sensing technology with landscape ecological theory has led to the development and use of a host of landscape metrics, which are measures of land use and/or land cover interpreted from satellite imagery (Cushman and others 2008; Gustafson 1998; Hargis and others 1998; Li and Wu 2004; Riitters and others 1995). Metrics have been used to analyze the spatial patterns of landscape change over scales ranging from watersheds and landscapes to regions, nations, and the globe (Cumming and Vervier 2002; De Clercq and others 2006; Gulinck and others 2001; Kupfer 2006; Riitters and others 2000; Southworth and others 2002; Yu and Ng 2006). Pattern metrics have a demonstrated utility in assessing whether critical components and functions of forests are being maintained (Garcia-Gigorro and Saura 2005), and can therefore potentially be used to aid in

J. A. Cardille (✉) · T. Holland
Department of Geography, University of Montreal,
520 Chemin Cote Ste Catherine Outremont,
Montreal H2V 2B8, Quebec, Canada
e-mail: jeffrey.cardille@umontreal.ca

J. C. White · M. A. Wulder
Canadian Forestry Service—Pacific Forestry Centre,
Victoria, British Columbia, Canada

national reporting of forest condition and change. The development of land-cover data sets derived by satellite across continental extents, such as the National Land Cover Dataset (NLCD) in the United States (Homer and others 2007) and the Earth Observation for Sustainable Development of Forests (EOSD) in Canada (Wulder and others 2008a), enable a national exploration the biotic, abiotic, and human processes that control the composition and configuration of landscapes over large areas.

When used judiciously, landscape metrics have the potential to quantify and elucidate aspects of forest loss and fragmentation. Landscape pattern calculation software such as FRAGSTATS (McGarigal and others 2002) and APACK (Mladenoff and Dezonia 1997) enable the calculation of a plethora of landscape pattern measurements, and extensive research has been aimed at identifying the key aspects of metrics intended to maximize their usefulness in various settings (Cain and others 1997; Cumming and Vervier 2002; Cushman and others 2008; De Clercq and others 2006; Li and others 2005; McAlpine and Eyre 2002; Riitters and others 1995). Nevertheless, there are important well-known caveats associated with the appropriate use of landscape metrics (Li and Wu 2004; McAlpine and Eyre 2002; Tischendorf 2001; Turner 2005), with no single metric suitable for all potential applications (Gergel 2007).

Depending on the spatial extent used for analysis, the calculation of landscape metrics across a large area can be computationally demanding. To enable processing and illustrate patterns across very large extents, regional- and continental-scale study areas are sometimes partitioned into smaller equal-sized analysis units (Cardille and others 2005; Cardille and Lambois 2010; Long and others 2010; Riitters and others 2004; Wulder and others 2008b). When a large country such as Canada or the United States is considered, there may be many thousands of analysis units to assess, each of which might contain hundreds of thousands or even millions of image pixels. In addition, the chosen land-cover classification scheme also influences both the complexity of metric calculations and subsequent applicability to a given management question.

Yet, even once large numbers of metrics are calculated for a large number of landscapes, conversion of a very large table of landscape metric values into a deep understanding of landscape patterns remains difficult. From the earliest development of landscape metrics to quantify spatial pattern, analysts have used principal components analysis or related data reduction techniques (e.g., Cushman and others 2008; Riitters and others 1995) to aid interpretation by compressing a large set of metrics to a statistically independent smaller set of meta-metrics. The intended effect is to summarize the metric values of a given set of landscapes by reducing the number of metrics that need to be interpreted.

Clustering algorithms provide the ability to compress a data set by grouping individuals having similar characteristics. Recent developments from the world of computer science offer potential application to any number of areas in which the ability to produce data outstrips the human ability to conceptualize it. The Affinity Propagation algorithm (Frey and Dueck 2007), in particular, finds optimal clusters substantially more efficiently than its rivals while simultaneously selecting a representative, or "exemplar", of each cluster identified in the set (Frey and Dueck 2007; Mezard 2007). This ability to highlight a small set of exemplars taken directly from a data set offers a valuable tool for several aspects of environmental management (Cardille and Lambois 2010). As a complement to existing tables or graphics of numerical characteristics, an objective set of representatives that encompasses the broad characteristics of a larger set can help managers to quickly and efficiently understand much about its contents. Additionally, because sets are summarized by the selection of representatives rather than by listing cluster attributes, exemplars can be directly inspected for an efficient understanding of the variety of the land-cover patterns in the larger set.

This research illustrates two applications that can be informed by this exemplar-guided approach for understanding land cover in forested Canada: one descriptive scenario, and one hypothetical management scenario. For the first application, we determine representative landscapes in each of Canada's forested ecozones. In the second application, we identify representative landscapes of Ontario's parks and protected areas, and then use them to locate similar landscapes in the province that currently do not enjoy protected status. This work extends and deepens the effort of Cardille and Lambois (2010) in two main ways. By using a much smaller set of landscape metrics (nine vs. 92), we dramatically reduce the amount of information available for distinguishing landscapes, providing a substantially greater challenge for the clustering algorithm. Second, by using identified exemplar landscapes as a key for comparing protected and non-protected areas, we illustrate how this work may be applied to the management and evaluation of Canada's vast forest resource.

Material and Methods

Study Area

Canada contains 10% of the world's forests and 30% of the world's boreal forests; these forests contribute $28.1 billion to the national balance of trade and provide an estimated 361,300 direct jobs annually (Natural Resources Canada 2008). Canada's forests support 180 different native

species of trees and provide habitat for more than 93,000 species of plants, animals, and micro-organisms (Natural Resources Canada 2008). Less than 1% of Canada's forests are harvested annually (Natural Resources Canada 2008).

Canada has 15 terrestrial ecozones, high-level divisions of the land mass according to climatic and vegetation patterns (Ecological Stratification Working Group 1995; Marshall and Schut 1999). Of these, ten are considered forested, vary substantially in size, and contain a range of forest ecosystem types. The ten forested ecozones occupy approximately 650 million ha (Wulder and others 2008b) and contain over 402 million ha of non-contiguous forests and other wooded land (Power and Gillis 2006). Ecozones offer a linkage to other national reporting activities while providing ecologically meaningful context (Bailey and others 1985; McMahon and others 2004). Despite the grouping of forested lands into ecozones, the land-cover patterns found within a given ecozone are not homogeneous, with descriptions of the vegetation in these zones detailed in Wulder and others (2008a). On the west coast of Canada, rugged topography and an influx of warm, moist air from the Pacific Ocean has resulted in a diverse and highly productive range of forest types. Subalpine forests are found in the mountainous areas of British Columbia and Alberta, while montane forests dominate the drier plateaus of central and southern British Columbia. In the central part of Canada, the Boreal Forest, primarily composed of coniferous species, stretches in a continuous belt from the Rocky Mountains (south) and Alaska (north) eastward to Newfoundland and Labrador, while deciduous species dominate in southern Ontario and Quebec (Rowe 1972).

Data

Land Cover and Fragmentation Metrics

The EOSD land cover product (hereafter referred to as EOSD LC 2000) was generated using circa 2000 Landsat satellite imagery to map 23 unique land cover classes in the forested ecozones of Canada (Wulder and others 2008a). The EOSD LC 2000 has a spatial resolution of 25 m, with approximately 10 billion 25 m pixels found within the forested ecozones of Canada. The 23 land cover classes were reclassified to forest, non-forest, or other to focus on the distribution and configuration of forest patterns (Wulder and others 2008b).

Using the reclassified EOSD LC 2000 product (Fig. 1), Wulder and others (2008b) selected nine key metrics to communicate the fragmentation trends present over Canada's forests: (1) proportion of area forested; (2) number of forest patches; (3) proportion of patches that are forested; (4) mean forest patch size; (5) forest patch size standard deviation; (6) amount of forest edge; (7) forest edge density; (8) forest/forest joint count; and (9) forest/non-forest join count. These metrics were selected because they

Fig. 1 Distribution of EOSD forest (*green*), nonforest (*yellow*) and other (*white*) classes used for this study
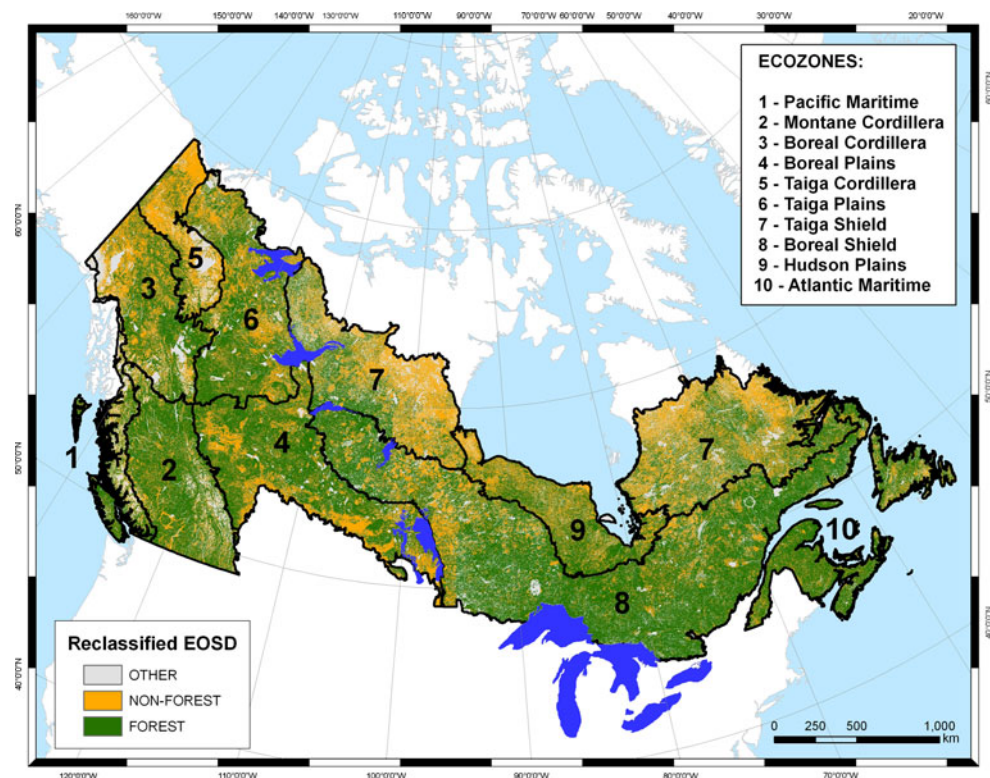
"depicted fragmentation as a condition of the landscape; captured the different types of fragmentation, as caused by natural and anthropogenic disturbances, ecosystem characteristics, and land use activities; were minimally redundant; and were readily interpretable and easy to understand when reported nationally" (Wulder and others 2008b). These were computed with freely available APACK software (Mladenoff and Dezonia 1997). Metrics were calculated for each of 7794 1:50,000 NTS map sheets (hereafter referred to as "landscapes"), which were used as analysis units in this study.

Application 1: Identifying Representative Landscapes of the Forested Ecozones of Canada

Although ecozones are a standard reporting unit for national-scale studies in Canada, each ecozone covers a very large area and cannot realistically be thought of as containing homogeneous land cover patterns. Because landscape patterns contain both aspects of landscape composition and aspects of landscape configuration, this heterogeneity is difficult to express with just a few numbers in tabular form. Even when only a few metrics are of interest, understanding and adequately expressing variability among hundreds or thousands of landscapes can be daunting. The typical approach is to select and summarize land-cover proportions or pattern metric values across given established reporting units (cf. Wulder and others 2008b). Such tables, while clearly informative and useful, summarize a vast area with set of numbers, which can be difficult to interpret and may not adequately express the remarkable variety that might lie within a given reporting unit. How might the information be expressed differently? In particular, might one usefully illustrate that variety using a small list of objectively chosen representative landscapes?

To locate these representatives objectively, we used the affinity propagation algorithm in order to group landscapes and simultaneously identify, for each cluster, a single member that best represents it (Frey and Dueck 2007). To find a set of exemplars, the algorithm operates on estimates of the similarity between pairs of objects; in this context, that meant estimating the similarity between all pairs of 1:50,000 map sheets in a given ecozone. Estimating similarities between land-cover patterns in a given ecozone was a multi-step process. First, because many pattern metrics are correlated (Cushman and others 2008; Riitters and others 1995), we performed a principal components analysis (PCA) of the nine selected metrics, scaling and rotating them to concentrate the maximum variation among metric values onto a set of orthogonal axes. The PCAs indicated that three independent axes existed among these metrics for most of the ecozones, with three eigenvalues

above or very near one. The three axes together captured between 85 and 95% of the variation among the nine metrics chosen for consideration among the landscapes of each ecozone. We used values along the PCA axes in order to ensure that calculations of similarity between landscapes were not biased toward certain aspects of landscapes that had been computed redundantly in the initial metric set. Using the principal component values, we estimated the similarity between the patterns in any pair of landscapes as the negative Euclidean distance between their principal component values (Frey and Dueck 2007). The resulting pairwise matrix represented our best estimate of the similarity in land cover composition and configuration between all pairs of landscapes in each ecozone. This allowed us to quickly request, for any ecozone, any number of clusters and, with them, the representative landscapes. Because there is no strictly correct number of clusters inherent in a set of data, the affinity propagation algorithm places no limit on the number of exemplars that can be identified. To illustrate the potential of this approach while keeping the total number of exemplars to discuss moderately small, we tasked the affinity propagation algorithm with identifying two representative landscapes for each of the ten ecozones.

Figure 2 shows the landscape exemplars for each forested ecozone. The clustering and choice of representatives provided by the affinity propagation algorithm were broadly consistent with known ecozone characteristics. Below, we describe the exemplar landscapes that the affinity propagation algorithm selected, ecozone by ecozone.

1. Boreal Shield (2,166 landscapes). The Boreal Shield ecozone is the largest terrestrial ecozone in Canada and is characterized by dense stands of conifers (white and black spruce, balsam fir, and tamarack) juxtaposed against communities of lichens, shrubs, forbs, and wetlands in areas of exposed bedrock and thinner soils. Exemplars were selected for two different landscapes: one dominated by forest (052O13) and the other dominated by a more complex mosaic of forest and non-forest (063O16); exemplar 052O13 is 78% forest, compared to 57% for exemplar 063O16, which also has 33% of its area occupied by wetland (see Figs. 2, 3 for this and the other ecozones).

2. Taiga Shield (1495 landscapes). This ecozone is characterized by two large biophysical features: the Taiga Forest and the Canadian Shield. The vegetation pattern in the ecozone is described as one of "innumerable lakes, wetlands, and open forests" (Ecological Stratification Working Group 1995). One exemplar is dominated by vegetated non-forest (024I05); the other, by forest (033O12).

Boreal Shield	063O16 (n = 965)			052O13 (n = 1201)
Taiga Shield	024I05 (n = 627)			033O12 (n = 868)
Pacific Maritime	093D07 (n = 106)			092K12 (n = 75)
Montane Cordillera	093O01 (n = 239)			093O10 (n = 282)
Boreal Cordillera	106D02 (n = 422)			095C10 (n = 271)
Taiga Cordillera	106C15 (n = 200)			116I03 (n = 175)
Taiga Plains	106J16 (n = 428)			085C05 (n = 372)
Boreal Plains	084J10 (n = 536)			083M02 (n = 406)
Hudson Plains	054A07 (n = 259)			053I08 (n = 158)
Atlantic Maritime	011E04 (n = 122)			021N04 (n = 82)

Fig. 2 Exemplar landscapes of each ecozone of forested Canada. Legend: *Green* forest; *Yellow* nonforest; *Black* water, ice, cloud, or cloud shadow. For each ecozone, the two exemplars selected by the affinity propagation algorithm of Frey and Dueck (2007) are shown. The number of 1:50,000 landscapes represented by each exemplar is given, as is the 1:50,000 map sheet index number of the landscape selected by the algorithm

3. **Pacific Maritime (181 landscapes).** This ecozone is dominated by temperate coastal forests composed of mixtures of western red cedar, yellow cedar, western hemlock, douglas fir, amabilis fir, mountain hemlock, sitka spruce, and alder. Both exemplars of the ecozone are dominated by forest, but reflect its mountainous topography, with one of the exemplars (093D07) representing higher-elevation forest and containing a greater proportion of non-vegetated cover (i.e., rock/rubble and exposed land). The second exemplar (092K12) represents low-elevation forest with little non-vegetated cover.

4. **Montane Cordillera (521 landscapes).** This ecozone is considered the most diverse in Canada, ranging from alpine tundra and dense conifer forests to dry sagebrush and grasslands (Ecological Stratification Working Group 1995). The two exemplars are both dominated by forest, but differ in the amount and arrangement of non-vegetated cover.

5. **Boreal Cordillera (693 landscapes).** Characterized by either closed or open forests at lower elevations and alpine tundra at higher elevations, this ecozone has mountain ranges with extensive plateaus. The less forested exemplar contains a greater amount of vegetated non-forest (i.e., shrub and herb) and non-vegetated (exposed land) cover. In the other exemplar

(095C10), valley-bottom forest dominates, mixed with higher-elevation shrub and alpine areas.

6. **Taiga Cordillera (375 landscapes).** This ecozone contains Canada's "largest waterfalls, deepest canyons, and wildest rivers" (Ecological Stratification Working Group 1995). The two exemplars are markedly different, with one (106C15) predominantly non-vegetated (i.e., rock/rubble and exposed land) with shrub, herb, and pockets of forest, and the other (116I03) a mix of forest and non-forest (i.e., shrub, herb).

7. **Taiga Plains (800 landscapes).** Slow-growing conifer forests of black spruce are the dominant vegetation in this ecozone. This is captured in the exemplar that is predominantly a mix of forest, wetland, and water (085C05). The other exemplar (106J16) has a heterogeneous distribution of cover types, including non-vegetated (water, exposed land), non-forest (bryoid, shrub, wetland), and forest (coniferous and deciduous).

8. **Boreal Plains (942 landscapes).** This ecozone is dominated by forest, primarily coniferous (black and white spruce, jack pine, and tamarack), with broadleaf forests found in transitional areas with prairie grasslands. Exemplar 084J10 captures a heterogeneous distribution of forest conditions, with coniferous, broadleaf, and mixed forests interspersed with shrub and wetland. Forest management is evident in some of the forest-dominated areas. The other exemplar (083M02), is dominated by herbs, consisting of agricultural lands in a regularized pattern.

9. **Hudson Plains (417 landscapes).** The Hudson Plains ecozone represents the largest extensive area of wetlands in the world (Ecological Stratification Working Group 1995) and the exemplars distinguish between wetland-dominated areas with forest and forest-dominated areas with wetlands.

10. **Atlantic Maritime (204 landscapes).** This ecozone is dominated by mixed stands of conifers and deciduous species. The exemplars reflect this pattern, with 011E04 being dominated by coniferous species and 021N04 dominated by mixed forests with pure coniferous and broadleaf stands as well. Both chosen representatives indicate areas that are subjected to forest harvesting and to agricultural uses.

This application indicated that a very small set of metrics, previously proposed as being useful for broad-scale reporting (Wulder and others 2008b), could successfully distinguish major landscape types from each other. Additionally, it revealed that the experimental technique of using affinity propagation could objectively identify exemplars that matched and informed existing, more subjective assessments of forested areas.
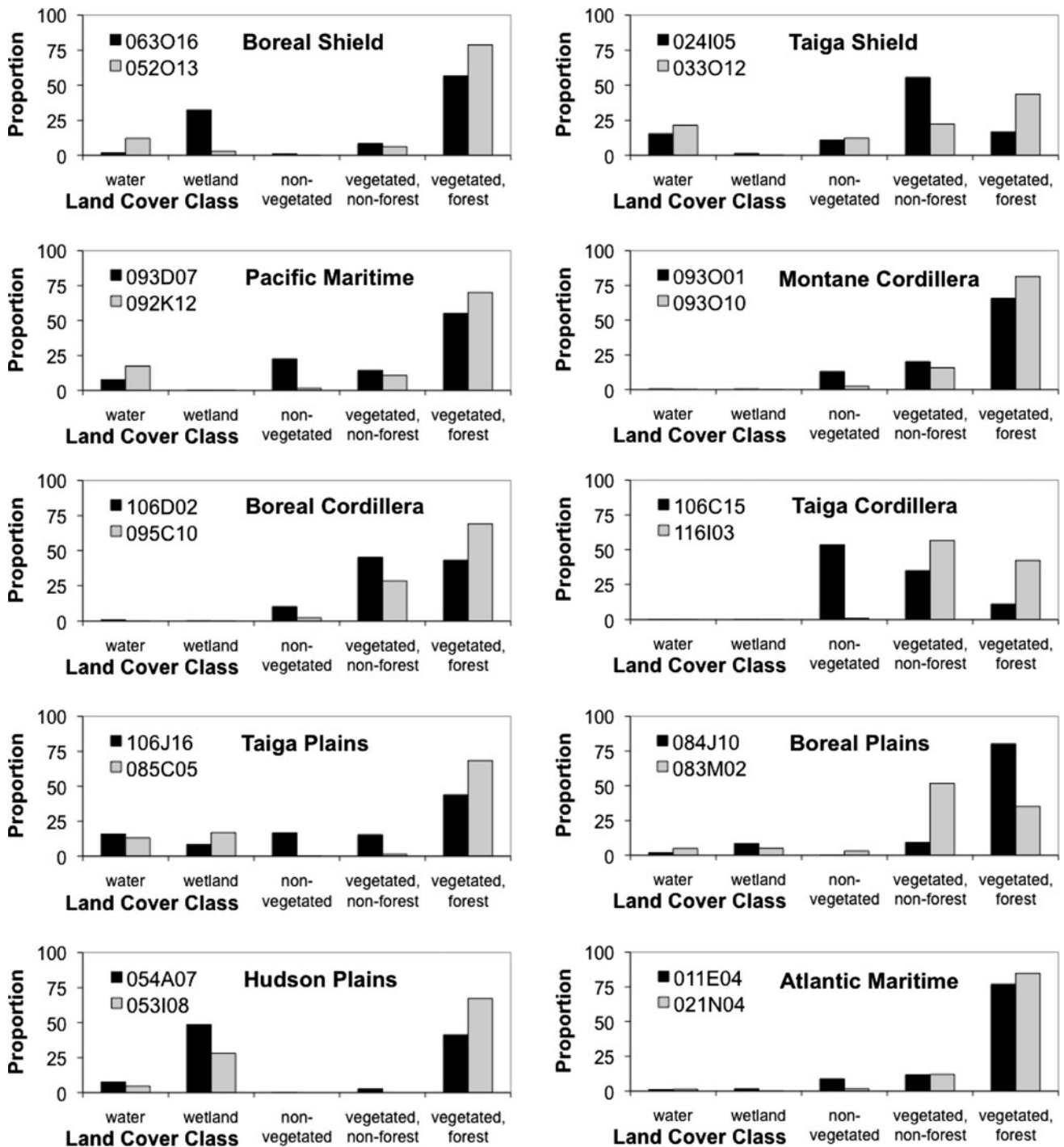
Fig. 3 Distributions of EOSD land cover classes within each of the exemplars

Application 2: Tracking Long-Term Landscape Change Inside and Outside Protected Areas in Ontario

For this application, we show how representative landscapes might be used to systematically track long-term landscape change inside and outside parks and protected areas (PPAs) in Ontario. We imagine a scenario in which

the provincial and/or federal governments have the resources to establish a limited number of long-term study areas to understand changes inside and outside the province's parks. Three PPA "flagship" landscapes will be studied in detail and, for each flagship, five landscapes will be identified outside the PPA system for long-term comparison. The fifteen landscapes will then be monitored and

analyzed (for example, for forest connectivity or fire frequency) as systematically identified near-replicates of the flagship landscapes. We use landscape metrics and affinity propagation (1) to find the three landscapes that best represent the land-cover variation inside protected areas; and (2) to identify, for each representative, the five landscapes outside protected areas that are the most similar, with respect to the identified landscape patterns.

Using a layer of parks and protected areas from the Government of Canada (http://geogratis.gc.ca/), we identified the 89 1:50,000 landscapes whose centers lay within the borders of a park or protected area in Ontario. We tasked the affinity propagation algorithm with identifying the three landscapes that best summarized the land-cover patterns of that set. (As with the previous application, the number of landscapes is chosen for parsimony and ease of presentation.) The PCA reduction of landscape metric values for the 89 PPA landscapes indicated three informative eigenvalues, and that that the first three principal components represented 61, 20, and 9% of the variance in the set, respectively. Asked to find the three best clusters and their exemplars, the algorithm identified the landscapes shown in Fig. 4 as representatives of the land-cover patterns in parks and protected areas of Ontario. The selected landscapes differ in several ways, most clearly in their proportion of forest cover.

For the second part of this hypothetical management application we identified, for each flagship landscape, its five most similar "relatives" outside the parks and protected

areas system. In this setting, these objectively chosen representatives would be used as study sites in which long-term observations could be contrasted with those in the flagship landscapes. We used the similarity values, as defined above, to estimate the similarity of land-cover patterns among all 918 landscapes of Ontario. We then used the resulting table of similarity values to identify the five landscapes whose metrics values were the nearest to those of the exemplars chosen in the first stage of this application. The related landscapes are visually quite similar to their target exemplars (Fig. 4). As we would hope, the five landscapes that were considered highly similar to a particular exemplar have properties that are both similar to each other and different from landscapes in other clusters. Taken as a whole, the consistency among these results indicates that several important facets of this approach function well. First, the landscape metrics of Wulder and others (2008b) were appropriate and sufficient for quantifying identifiable similarities in spatial patterns among landscapes. Second, the compression of these landscape metric values into a single similarity measure between landscapes retains information suitable for identifying landscapes with land-cover patterns that are similar in appearance.

Discussion

A national set of landscape metrics and the ability to identify a representative set of landscape patterns present
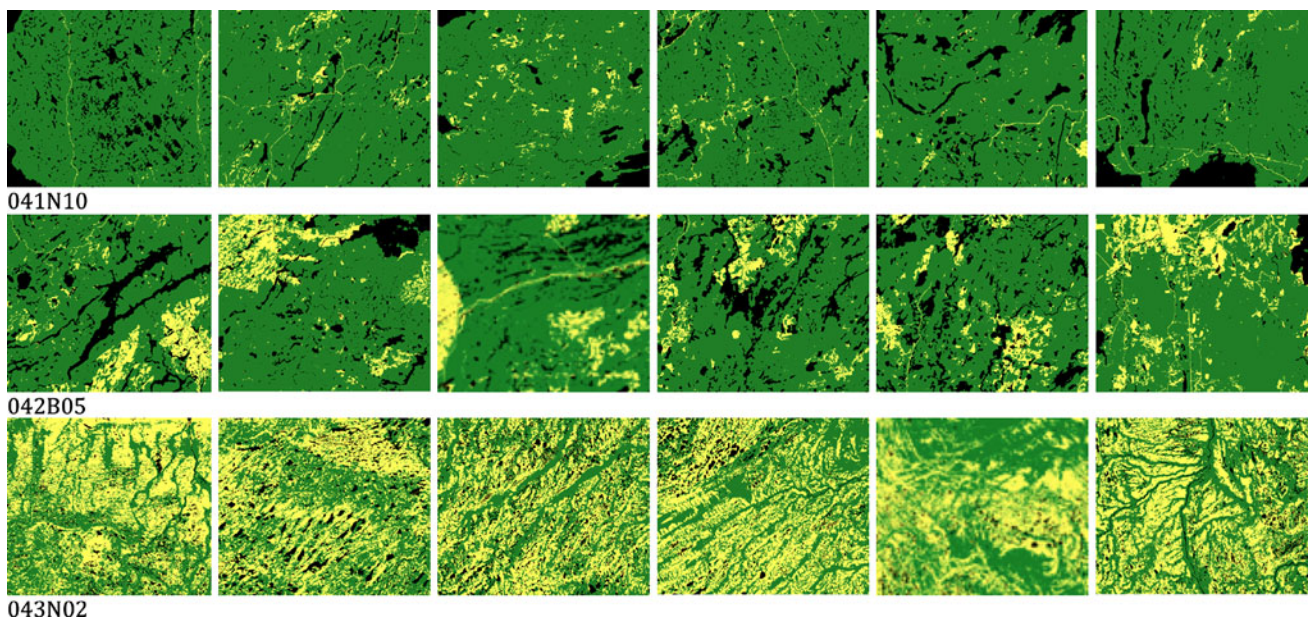
Fig. 4 Exemplar landscapes inside (*first column*) and related landscapes (*remaining columns*) outside of the parks and protected areas in Ontario, Canada. To the right of each exemplar are the five landscapes outside the parks and protected areas system that have the most similar landscape patterns to the representatives, as estimated using landscape metric values. The resemblance of an exemplar to the landscapes it represents is an indicator of the success with which the metrics and algorithm can successfully relate the landscapes for such applications. Legend: *Green* forest; *Yellow* nonforest; *Black* water, ice, cloud, or cloud shadow

the opportunity to better understand the variation in landscape pattern, its distribution in the ecozones of forested Canada, and its potential use in management scenarios across large areas.

In the first application presented here, the selected exemplars for each ecozone provide critical context that complements a more generalized presentation of ecozone characteristics. Wulder and others (2008b) presented a summary of fragmentation patterns within each ecozone, providing a broad overview of regional differences in landscape patterns; however, such a tabular summary cannot easily convey the variability and full range of landscapes that are to be found within each ecozone. For example, the Boreal Plains ecozone is described in Wulder and others (2008b) as being approximately 62 ± 25% forested. Results of the approach here add greatly to that information, with its choice of exemplars for the Boreal Plains ecozone that are clearly quite different: one that is dominated by forest (084J10; 81% forest) and one that is dominated by vegetated non-forest (083M02, 38% forest). As mentioned previously, a user wishing to observe more of the variability could simply re-task the clustering algorithm to generate more representatives. We found the ability to inspect images of individual exemplars to be a welcome addition to other established ways of conveying landscape characteristics across these vast areas.

As described in the Application 2 section, the systematic identification of representative landscapes might be used as part of a conservation and/or monitoring strategy in several ways by federal and provincial natural resource managers. Because selected exemplars represent other landscapes that have similar forest composition and configuration, they would be ideal candidates for guiding the establishment of permanent sample plots or long-term study sites. Second, because the process of exemplar selection provides objective measures of the similarity of all pairs of landscapes, natural resource managers could use this approach to help assemble sets of similar landscapes in which to study the effects of various management strategies. Landscapes that are similar to flagship protected areas, but which themselves are not protected, could serve either as systematically identified replicates for future studies or investigated further for conservation purposes, perhaps to increase the resilience of the overall conservation portfolio.

There are several important aspects of this approach that should be carefully considered before undertaking a management implementation of these ideas. (1) The visual evaluation of representatives was sufficient to illustrate these concepts; in contrast, any real-world use of representatives for long-term study would need to be preceded by extensive evaluation of their characteristics. This might certainly include the consideration of many other factors beyond land-cover metric values, including consideration

of the political, social, and financial characteristics of the landscapes in question. (2) The suite of nine fragmentation metrics were chosen for their suitability for national reporting and their ease of calculation and interpretation (Wulder and others 2008b). This choice resulted in representative landscapes with respect to the chosen metrics. It does not mean they identified similarities or differences that were universally relevant to a particular management decision, and other applications seeking representative landscapes might well demand the use of other metrics. Furthermore, since landscape metrics are typically sensitive to changes in classification scheme and extent, exemplars emerging based on analysis of other land cover products may differ (Gergel 2007). (3) In the Application 1 section, we identified two landscapes to represent each of the ten ecozones, primarily to keep the discussion tractable. In the Application 2 section, we identified three park landscapes as flagships. In our view, there is no inherently "correct" number of exemplars in a given ecozone for all applications. There is an extensive body of literature on extracting an optimal number of clusters from a set (e.g., Jain and others 1999): although there is some convergence around certain estimation techniques, there is no agreed-upon single estimator. For certain applications, the selection of a greater number of exemplars may be preferable, for example in proportion to the number of landscapes. The affinity propagation algorithm can select any number of exemplars specified by the user. (4) An additional factor worth considering is the composition of the similarity value for a given research question. In the applications presented here, all measured aspects of landscape pattern were weighted equally. For some applications, a variable weighting might be preferred, for example to emphasize the importance of edges in computing the similarities between landscapes. (5) Finally, it is worth noting that the EOSD LC 2000 classification was substantially simplified from its 23 native classes (Wulder and others 2008a) for the purposes of generating the fragmentation metrics (Wulder and others 2008b). Even with this greatly reduced information, the metrics enabled the affinity propagation algorithm to distinguish amongst different broad land cover assemblages.

This exemplar-guided approach using affinity propagation is extremely well-suited to these specific management questions. The commonly used PCA algorithm does reduce the dimensionality of the data used to quantify landscape patterns, and is used here for that purpose, but does not directly provide clear groupings with which to compare individual or groups of landscapes. Meanwhile, most well-known clustering algorithms, (e.g., *k*-means) are not designed to produce representative items, and the analysis of their results is best limited to numerical tabulations of summary group characteristics. In contrast to those

clustering strategies, affinity propagation is a member of a very small family of algorithms (*k*-medoids being the best-known other such algorithm) designed to both cluster data and select representatives from the data set.

In technical terms, this research adds to the small but growing body of literature concerned with performing useful segmentation of landscape patterns in large satellite-based data sets. Without further considerable research, however, one should be cautious in comparing the specific results of this study with other large-scale studies in a similar spirit. Efforts by Long and others (2010), while quite successful at clustering landscapes based on metric values, were of a much smaller landscape size (1 km^2 in that study vs. 800 km^2 here), covered a much smaller total area (5 million ha vs. 650 million ha here), and were based on a different satellite classification. The identification of exemplars for the continental United States (Cardille and Lambois 2010) used the same basic protocol as that used here; however, the underlying satellite classifications differed substantially between the EOSD (Wulder and others 2008a) and the NLCD (Vogelmann and others 2001). Most importantly, the NLCD classification represents land use as well as land cover. To the extent that underlying classifications can be made consistent across national borders, a project to identify exemplar landscapes of all North America might be fruitful.

Citations and applications of the affinity propagation algorithm in the broader literature suggest that it is experiencing widespread adoption in an extremely wide range of settings. In our view, the flexibility and simplicity of the method could be widely useful in environmental management. As awareness of and access to large databases of landscape characteristics increase, managers might draw inspiration from affinity propagation's growing use in other fields, such as to extract representative image bands from data-heavy hyperspectral remote-sensing data (Qian and others 2009), to cluster large numbers of online videos by selecting representative image frames (Karpenko and Aarabi 2011), to tag photo albums (Liu and others 2011), or to identify temporal features in gene expression data (Kiddle and others 2010).

Although this clustering of landscapes was built using the nine landscape metrics of Wulder and others (2008b), the approach outlined here is not limited to these metrics nor, indeed, to land cover-based metrics in particular. Because the algorithm operates on any calculated similarities among landscapes, one could readily select representative landscapes based on other criteria that can be measured or estimated in each unit. For example, managers interested in understanding the effects of insect outbreaks could combine measures of pre-infestation conditions, attack intensity, attack duration, and post-attack conditions into a similarity measure between landscapes. Then, any number of representatives could be identified for locating study sites that most efficiently represent areas having similar attack histories and recovery trajectories. The algorithm is not limited to landscapes of a certain size, nor of an equal size: one might identify representative forest management units, for example by combining estimates of the percentage of forest harvested in each decade of the 20th century into a similarity measure. Each identified forest management unit exemplar would then represent a large number of other units having similar management histories. Similarly, managers interested in biodiversity could combine presence/absence measures of an arbitrarily large number of species into a measure of similarity, to cluster and identify individual landscapes that support similar combinations of species. Moreover, the clusters of management units having similar characteristics could potentially be used to ensure that the locations of future harvests are chosen from clusters of management units having desired criteria. Measures contributing to calculations of similarity need not be inherently numeric in nature; measures like the Adjusted Rand Index (Hubert and Arabie 1985) can quantify similarity among sets of nominal variables. This ability to quickly and objectively identify similar landscapes might be especially useful for exploring either "in-kind" or "out-of-kind" development offsets for conservation planning (Kiesecker and others 2010). Even more generally, the affinity propagation algorithm is not limited to landscapes; as a generic clustering and classification tool, it could be used to identify, for example, representative lakes from a larger set of lakes given specified criteria of interest.

Conclusion

As more large and complex land cover data sets such as the EOSD LC 2000 are produced, it will be increasingly important to have flexible strategies to understand them. Because land-cover datasets like the EOSD are rooted in human perceptions of the environment, the identification of viewable representatives is perhaps an especially informative approach for developing an understanding of Canada's vast forest resource. The landscapes described and presented herein are useful in several specific ways. First, because the analysis unit is at the same 1:50,000 scale as the Government of Canada's National Topographic map series, the exemplars have a scale and appearance that is familiar to many natural resource managers and ecologists. As shown in the Application 1 section, the ability to view representative landscapes is a powerful supplement to strictly tabular or statistical summaries of landscape patterns. Second, the numerical estimates of similarity amongst landscapes that are provided by the affinity

propagation algorithm can help illustrate and identify connections among landscapes, as shown in the Application 2 section.

Landscape metrics, when properly and conservatively used, provide the ability to make objective comparisons among different assemblages of land cover. In large data sets, data compression to date has mainly been accomplished through reduction techniques operating on redundancy in landscape metrics: through this compression, the number of metrics to consider in a study is greatly reduced. But in national-scale studies that encompass even a few metrics or meta-metrics, the vast number of landscapes remains, with differences among them remaining difficult to interpret. By grouping landscapes into related sets and highlighting an optimal representative of each group, the affinity propagation approach appears to offer benefits for analysts wishing to extract meaningful information from large landscape data sets. In addition to its potential benefit in helping to set management priorities across large areas, the identification of a single landscape as an exemplar is a benefit that opens the interpretation of landscape patterns to a much larger group of both scientists and laypeople.

References

Bailey R, Zoltai S, Wiken E (1985) Ecological regionalization in Canada and the United States. Geoforum 16:265–275

Cain DH, Riitters K, Orvis K (1997) A multi-scale analysis of landscape statistics. Landscape Ecology 12:199–212

Cardille JA, Lambois M (2010) From the redwood forest to the Gulf stream waters: human signature nearly ubiquitous in representative US landscapes. Frontiers in Ecology and the Environment 8:130–134

Cardille J, Turner M, Clayton M, Gergel S, Price S (2005) METALAND: characterizing spatial patterns and statistical context of landscape metrics. Bioscience 55:983–988

Cumming S, Vervier P (2002) Statistical models of landscape pattern metrics, with applications to regional scale dynamic forest simulations. Landscape Ecology 17:433–444

Cushman SA, McGarigal K, Neel MC (2008) Parsimony in landscape metrics: strength, universality, and consistency. Ecological Indicators 8:691–703

De Clercq EM, Vandemoortele F, De Wulf RR (2006) A method for the selection of relevant pattern indices for monitoring of spatial forest cover pattern at a regional scale. International Journal of Applied Earth Observation and Geoinformation 8:113–125

Ecological Stratification Working Group (1995) A national ecological framework for Canada. Agriculture and Agri-Food Canada, Research Branch, Centre for Land and Biological Resources Research, and Environment Canada, State of the Environment Directorate, Ecozone Analysis Branch, Ottawa/Hull

Frey BJ, Dueck D (2007) Clustering by passing messages between data points. Science 315:972–976

Garcia-Gigorro S, Saura S (2005) Forest fragmentation estimated from remotely-sensed data: is comparison across scales possible? Forest Science 51:51–63

Gergel S (2007) New directions in landscape pattern analysis and linkages with remote sensing. In: Wulder MA, Franklin SE (eds) Understanding forest disturbance and spatial pattern: remote sensing and GIS approaches. CRC Taylor & Francis, Boca Raton, pp 173–208

Gulinck H, Mugica M, de Lucio JV, Atauri JA (2001) A framework for comparative landscape analysis and evaluation based on land cover data, with an application in the Madrid region (Spain). Landscape and Urban Planning 55:257–270

Gustafson EJ (1998) Quantifying landscape spatial pattern: what is the state of the art? Ecosystems 1:143–156

Hargis CD, Bissonette JA, David JL (1998) The behavior of landscape metrics commonly used in the study of habitat fragmentation. Landscape Ecology 13:167–186

Homer C, Dewitz J, Fry J, Coan M, Hossain N, Larson C, Herold N, McKerrow A, VanDriel JN, Wickham J (2007) Completion of the 2001 national land cover database for the conterminous United States. Photogrammetric Engineering & Remote Sensing 73:337–341

Hubert L, Arabie P (1985) Comparing partitions. Journal of Classification 2:193–218

Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. Acm Computing Surveys 31:264–323

Karpenko A, Aarabi P (2011) Tiny videos: a large data set for nonparametric video retrieval and frame classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 33:618–630

Kiddle SJ, Windram OPF, McHattie S, Mead A, Beynon J, Buchanan-Wollaston V, Denby KJ, Mukherjee S (2010) Temporal clustering by affinity propagation reveals transcriptional modules in *Arabidopsis thaliana*. Bioinformatics 26:355–362

Kiesecker J, Copeland H, Pocewicz A, McKenney B (2010) Development by design: blending landscape-level planning with the mitigation hierarchy. Frontiers in Ecology and the Environment 8:261–266

Kupfer JA (2006) National assessments of forest fragmentation in the US. Global Environmental Change 16:73–82

Li HB, Wu JG (2004) Use and misuse of landscape indices. Landscape Ecology 19:389–399

Li X, He HS, Bu R, Wen Q, Chang Y, Hu Y, Li Y (2005) The adequacy of different landscape metrics for various landscape patterns. Pattern Recognition 38:2626–2638

Liu D, Wang M, Hua XS, Zhang HJ (2011) Semi-automatic tagging of photo albums via exemplar selection and tag inference. IEEE Transactions on Multimedia 13:82–91

Long J, Nelson T, Wulder M (2010) Regionalization of landscape pattern indices using multivariate cluster analysis. Environmental Management 46:134–142

Marshall I, Schut P (1999) A national ecological framework for Canada. Agriculture and Agrifood Canada. http://sis.agr.gc.ca/cansis/nsdb/ecostrat/intro.html. Accessed 24 August 2010

McAlpine CA, Eyre TJ (2002) Testing landscape metrics as indicators of habitat loss and fragmentation in continuous eucalypt forests (Queensland, Australia). Landscape Ecology 17:711–728

McGarigal K, Cushman SA, Neel MC, Ene E (2002) FRAGSTATS: spatial pattern analysis program for categorical maps. www.umass.edu/landeco/research/fragstats/fragstats.html. Accessed 24 August 2010

McMahon G, Wiken EB, Gauthier DA (2004) Toward a scientifically rigorous basis for developing mapped ecological regions. Environmental Management 34:S111–S124

Mezard M (2007) Where are the exemplars? Science 315:949–951

Mladenoff DJ, Dezonia B (1997) APACK 2.0 user's guide. University of Wisconsin, Madison

Natural Resources Canada (2008) The State of Canada's forests: annual report 2008. Natural Resource Canada, Canadian Forest Service, Ottawa

Power K, Gillis M (2006) Canada's forest inventory 2001. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria

Qian Y, Yao F, Jia S (2009) Band selection for hyperspectral imagery using affinity propagation. IET Computer Vision 3:213–222

Riitters KH, O'Neill RV, Hunsaker CT, Wickham JD, Yankee DH, Timmins SP, Jones KB, Jackson BL (1995) A factor-analysis of landscape pattern and structure metrics. Landscape Ecology 10:23–39

Riitters KH, Wickham JD, O'Neill RV, Jones KB, Smith ER (2000) Global-scale patterns of forest fragmentation. Conservation Ecology 4:3

Riitters K, Wickham J, Coulston J (2004) A preliminary assessment of Montreal process indicators of forest fragmentation for the United States. Environmental Monitoring and Assessment 91:257–276

Rowe JS (1972) Forest regions of Canada. Information Canada, Ottawa, 172 p, AMICUS No. 2307881

Southworth J, Nagendra H, Tucker C (2002) Fragmentation of a landscape: incorporating landscape metrics into satellite analyses of land-cover change. Landscape Research 27:253–269

Tischendorf L (2001) Can landscape indices predict ecological processes consistently? Landscape Ecology 16:235–254

Turner M (2005) Landscape ecology: what is the state of the science? Annual Review of Ecology Evolution and Systematics 36:319–344

Vogelmann JE, Howard SM, Yang LM, Larson CR, Wylie BK, 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 and Remote Sensing 67:650–662

Wulder M, White J, Cranny M, Hall R, Luther J, Beaudoin A, Goodenough D, Dechka J (2008a) Monitoring Canada's forests—part 1: completion of the EOSD land cover project. Canadian Journal of Remote Sensing 34:548–549

Wulder MA, White J, Han T, Coops N, Cardille J, Holland T, Grills D (2008b) Monitoring Canada's forests—part 2: national forest fragmentation and pattern. Canadian Journal of Remote Sensing 34:563–584

Yu X, Ng C (2006) An integrated evaluation of landscape change using remote sensing and landscape metrics: a case study of Panyu, Guangzhou. International Journal of Remote Sensing 27:1075–1092