Landscape Ecol (2008) 23:505–511
DOI 10.1007/s10980-008-9215-x

PERSPECTIVE

On the accuracy of landscape pattern analysis using remote sensing data

Guofan Shao · Jianguo Wu

Received: 21 February 2008 / Accepted: 26 February 2008 / Published online: 14 March 2008

Abstract Advances in remote sensing technologies have provided practical means for land use and land cover mapping which is critically important for landscape ecological studies. However, all classifications of remote sensing data are subject to different kinds of errors, and these errors can be carried over or propagated in subsequent landscape pattern analysis. When these uncertainties go unreported, as they do commonly in the literature, they become hidden errors. While this is apparently an important issue in the study of landscapes from either a biophysical or socioeconomic perspective, limited progress has been made in resolving this problem. Here we discuss how errors of mapped data can affect landscape metrics and possible strategies which can help improve the reliability of landscape pattern analysis.

G. Shao (✉)
Department of Forestry and Natural Resources, Purdue University, 715 West State Street, West Lafayette, IN 47907-1159, USA
e-mail: shao@purdue.edu

J. Wu
School of Life Sciences & Global Institute of Sustainability, Arizona State University, Tempe, AZ 85287-4501, USA

J. Wu
Sino-US Center for Conservation, Energy and Sustainability Science (SUCCESS), Inner Mongolia University, Hohhot 010021, China

Keywords Landscape pattern · Land use and land cover maps · Classification accuracy · Landscape metrics · Errors · Remote sensing

Introduction

One of the primary goals of landscape ecology is to elucidate the relationship between spatial pattern and ecological processes (Turner 2005; Wu and Hobbs 2002). Accurately quantifying landscape pattern is a prerequisite for achieving this goal. While a large number of metrics have been developed for quantifying landscape pattern since the seminal paper by O'Neill et al. (1988), various conceptual and technical problems still exist. For example, Li and Wu (2004, 2007) recently discussed three kinds of such problems: conceptual flaws in landscape pattern analysis, inherent limitations of landscape indices, and improper use of landscape indices, and they proposed a series of guidelines for overcoming these problems to assure the effective application of landscape pattern analysis methods.

Because most studies of landscape pattern analysis use classified thematic maps based on remote sensing data, the accuracy or uncertainty associated with the maps is critical for our ability to reliably characterize spatial pattern, detect changes, and relate pattern to process (Hess 1994; Hess and Bay 1997; Wu and

Hobbs 2002; Iverson 2007). Unfortunately, all remote sensing-based classifications are subject to different kinds of errors, and these errors will be carried over or even propagated in subsequent landscape pattern analysis. Without knowing the magnitude of errors or uncertainties in landscape data, the characterization of landscape pattern is hardly reliable, inferences on pattern–process relationships may be flawed, and consequently, recommendations for conservation and management are unwarranted. Nevertheless, limited progress has been made in explicitly examining the impacts of classification errors in spatial data on landscape pattern analysis (but see Wickham et al. 1997; Fang et al. 2006). Rigorously addressing this problem is one of the research priorities in landscape ecology today (Wu and Hobbs 2002; Iverson 2007), and has important implications for ecology in general (Wu et al. 2006).

The applications of remote sensing data in landscape ecological studies are pervasive (Groom et al. 2006; Iverson 2007). With the rapidly increasing availability of remotely sensed data and continued widespread use of landscape pattern metrics, the problem of classification accuracy may affect many biophysical and socioeconomic studies of landscapes. The objectives of this paper, therefore, are to discuss how the inaccuracy of image data classification can affect landscape pattern analysis and to explore several possible approaches to improving the reliability of landscape metrics.

Classification accuracy and reliability of landscape metrics

The accuracy of land use and land cover (LULC) maps that are derived with digital remote sensing data is usually represented in terms of producer's accuracy, user's accuracy, and overall accuracy, which are commonly calculated from an error matrix (or confusion matrix; see Congalton and Green 1999). For clarity and consistence of the discussion, we briefly present the definitions of these terms as commonly used in the remote sensing community (see http://ccrs.nrcan.gc.ca/; http://landcover.usgs.gov/accuracy/). Producer's accuracy is the percentage of a particular LULC type on the ground is correctly classified in the map, and measures the error of omission (1 − producer's accuracy). It is calculated as the ratio of the number of correctly classified pixels for a class (i.e., LULC type) to the total number of ground truth pixels for that class. User's accuracy is the percentage of a class on the map that matches the corresponding class on the ground, and measures the error of commission (1 − user's accuracy). Therefore, "producer's accuracy is a measure of the accuracy of a particular classification scheme," whereas "user's accuracy is a measure of the reliability of an output map generated from a classification scheme" (http://ccrs.nrcan.gc.ca/). While producer's and user's accuracy deal with individual classes, overall accuracy is the percentage of correctly classified pixels out of all pixels sampled for all classes.

There is a common assumption that the higher the classification accuracy, the more accurate a map product, and thus the more reliable landscape indices derived from the map. However, there are at least two reasons why this assumption needs to be taken cautiously. First, there is not a universally applicable standard based on which the adequacy of classification accuracy can be quantified, and most of the publicly accessible landscape data have a lower classification accuracy than usually expected. For example, the USGS standard of image data classification was originally set to 85% for overall accuracy a priori (Anderson et al. 1976), but the actual accuracy of image classification rarely reached that standard. The overall accuracy of the land cover map for the eastern United States derived from the 1992 Landsat TM data is 81% at Anderson level I (with land cover types of water, urban, barren land, forest, agricultural land, wetland, and rangeland), and decreases to 60% at Anderson level II (Vogelmann et al. 2001). The 17 class-IGBP (International Geosphere-Biosphere Programme) land cover map has an overall accuracy of 73.5% (Scepan 1999). Due to high costs involved in accuracy assessment, many remotely sensed map products have not been assessed for accuracy, and thus the accuracy of landscape metrics computed based on such remote sensing products is completely unknown. Unfortunately, because of the lack of a generally accepted threshold for classification accuracy, such blind use of mapped data has been widespread in landscape studies. Recent studies have shown that the accuracy of landscape indices often has an exponential relationship with classification accuracy (Fig. 1; see Shao et al. 2001; Li and Wu 2004; Shao and Wu 2004).

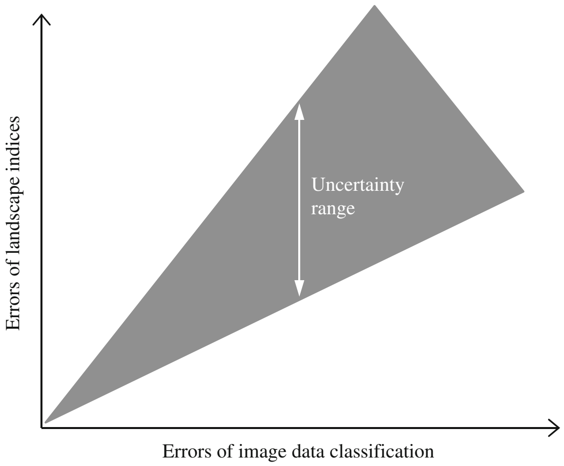

Fig. 1 An illustration of GIGO (gray in, gray out) effect using a semigarithmic scale on y-axis. Both the magnitude and uncertainty range of landscape index errors have an exponential relationship with the errors of image data classification

Second, LULC mapping with remote sensing data is hardly repeatable, and thus the same remote sensing imagery does not necessarily produce the same classified maps. Computer-aided automated or semi-automated approaches have become the dominant means in image classification as the quantity and variety of remote sensing data have increased exponentially in recent decades. Automated or semi-automated classification schemes often do not create maps of the same accuracy. For example, two global land cover data sets of 1-km resolution derived from 1992–1993 Advanced Very High Resolution Radiometer (AVHRR) imagery have been made available by the IGBP Data and Information System (IGBP-DIS) and University of Maryland—known as the DISCover and UMd 1 km LULC maps, respectively. Hansen and Reed (2000) showed that, because the two groups of analysts used different classification techniques, the global area totals of aggregated vegetation types had a per-pixel agreement of only 74% between the two maps. The individual class agreement between the two maps is only 49%. The DISCover map, in general, has more forests, whereas the UMd map has considerably more woody savanna/woodland and savanna/wooded grassland with an intermediate level of tree cover. If both datasets are used to characterize LULC pattern at a global or continent scale, the results and their interpretations are expected to differ significantly. A provocative question may follow: should we trust any of them?

Poor repeatability of image classification is also common for geospatial data at landscape and regional scales (Shao et al. 2001). For most remote sensing applications, an image-classification project only produces a single LULC map for a given area at a given time. Since there is no practical way to prove that this map is the best possible, there is no way to assure that landscape metrics derived from the map are the most accurate. Powell et al. (2004) discussed the problem of poor repeatability in image classification, and suggested that multiple interpreters are needed to work together in accuracy assessment. Groom et al. (2006) suggested to re-examine the roles of remote sensing in landscape ecology with regard to image data classification.

Suggestions for improving the reliability of landscape metrics

Given the uncertainties associated with many categorical maps derived from remote sensing data, how can we improve the accuracy of landscape analysis? Here we discuss several possible approaches that should be used in combination whenever possible.

Select appropriate classification techniques

Although the relationship between image classification accuracy and landscape metrics is often nonlinear, the variation in the values of landscape metrics becomes lower when classification accuracy is higher (Shao et al. 2001; Shao and Wu 2004). As a rule of thumb, most landscape metrics tend to stabilize in their values as the overall accuracy of image classification approaches 90%. This means that a high degree of classification accuracy is required for assuring the consistency and reliability of landscape metrics. Of course, the degree of the accuracy of landscape metrics is always dependent upon the specific objective of a study.

There are a number of advanced techniques available for improving classification accuracy. Commercial packages of image processing provide convenient interfaces that facilitate the use of common classification methods, but they may not incorporate the most rigorous and powerful classifiers better suited for landscape studies. For example, Kettig and Landgrebe (1976) proposed a classifier called ECHO (Extraction and Classification of Homogeneous Objects), in which the classification unit is conceptually similar to a patch as used in landscape ecology. Studies have shown that the ECHO classifier produces maps with much higher accuracy than those with the commonly used Maximum Likelihood classifier (Wu and Shao 2002; Lu et al. 2004). The object-oriented image analysis software, eCognition or Definiens, allows implementing expert knowledge, generates homogeneous objects through a local optimization procedure, and creates a hierarchical framework of decomposable image objects (Benz et al. 2004). The concepts and outcomes of eCognition applications are consistent with landscape pattern analysis. Several works have demonstrated the usefulness of eCognition in habitat mapping (e.g., Lathrop et al. 2006).

Select imagery of appropriate scales

Remote sensing data come with a variety of spatial and temporal resolutions. Although it may not always be feasible, matching the spatial and temporal resolutions of the imagery with the spatial and temporal grain sizes at which a landscape is intended to be characterized. Otherwise, a scale mismatch between data and analysis would occur, a problem closely related to the issue of scale effects that has been widely studied (e.g., Turner et al. 1989; Wickham and Riitters 1995; Jelinski and Wu 1996; Wu 2004). For example, a narrow creek may be readily detectable on a 30 m-resolution Landsat TM image, but not on a 80 m-resolution Landsat MSS image. If pixel sizes are relatively large and patches of interest are relatively small, measurements of patch shape becomes meaningless. Similar effects are found with temporal resolutions or grain sizes. Vegetation has different spectral responses in different seasons, and leaf-on and leaf-off are two distinct seasons for remote sensing of vegetation. Discrepancies in the timing and interval of remote sensing images may result in major uncertainties in comparing different landscapes or detecting changes of the same landscape.

In general, if remote sensing images have a much finer resolution than the intended grain size of analysis, one may either coarse-grain the images (degradation) and then compute landscape metrics, or compute landscape metrics first and then upscale the metrics to a coarser grain size. Either of these two approaches may be problematic. Image degradation not only loses information on structural details but also may distort landscape patterns (Moody and Woodcock 1995; Saura 2004). Upscaling landscape metrics across grain sizes is often difficult or impossible because most metrics show erratic and unpredictable scaling behavior (Wu 2004).

Balance producer's and user's accuracy

As mentioned above, classification accuracy can be expressed in terms of producer's accuracy and user's accuracy, which correspond to errors of omission and commission, respectively. In most landscape ecological applications, the area of each LULC type (or class area) may be the most important measure. Shao et al. (2003) derived an index called Relative Error of Area (*REA*) from the error matrix, in which the accuracy of class area on a map is a function of producer's and user's accuracy. That is, $REA_k = ((1/UA_k) - (1/PA_k)) \times 100$, where UA_k and PA_k are user's and producer's accuracy for class *k*, respectively. The area of class *k* is overestimated if $REA_k > 0$, and underestimated if $REA_k < 0$. Evidently, REA_k becomes smaller when producer's and user's accuracy are numerically more similar, or when the errors of omission and commission converge. When $REA_k = 0$, the areal estimate of class *k* is 100% accurate statistically although the spatial pattern of class *k* may not be perfectly accurate unless both UA_k and PA_k are perfect. Shao et al. (2003) demonstrated that the actual accuracy of areal estimates of LULC types is highly correlated with REA_k, but not consistently with UA_k, PA_k, or overall accuracy.

Because it is practically impossible to get 100% accuracy in UA_k and PA_k to achieve a REA_k of zero, balancing producer's and user's accuracy can help improve the accuracy of areal estimates of LULC types. This is particularly important for large-scale landscape studies and change detections. When UA_k and PA_k both are high and close to each other in value, the reliability of landscape metrics of even spatial configuration is expected to be high as well.

Be aware of landscape metrics sensitive to classification errors

The sensitivity of landscape metrics to changing scale of analysis is now fairly well understood (Shen et al.

2004; Wu 2004), but the sensitivity of many landscape metrics to classification accuracy is much less known. Nevertheless, several recent studies have shown that some landscape metrics (e.g., mean patch size and patch density) are more sensitive to classification accuracy than others (Hess and Bay 1997; Wickham et al. 1997; Shao et al. 2001; Li and Wu 2004). If a landscape metric is insensitive to differences in landscape pattern, it certainly reduces the variation of the metric caused by classification errors, but also fails to be able to detect landscape structural changes that may be important to understanding ecological processes. However, landscape metrics that are good at detecting pattern changes are more likely those also sensitive to classification errors. To resolve this problem, we must know the level of classification errors that can be tolerated for achieving an acceptable degree of reliability of landscape metrics. Wickham et al. (1997) examined the sensitivity of three landscape metrics (average patch compaction, contagion, and fractal dimension). These authors found that the misclassification rate needed to be at least 5% smaller than the actual differences in LULC composition in order to be confident that differences in landscape metrics were not due merely to classification errors. More comprehensive and systematic studies of this sort with different kinds of landscape patterns (i.e., ecotones) are needed to test and expand these important findings (e.g., Arnot et al. 2004).

Multi-factor assessment of accuracy

Because there is no universally acceptable approach to assessing the reliability of landscape indices, it seems important to consider multiple assessment criteria in landscape pattern analysis (Table 1). In this example, we have selected a list of assessment factors that are relevant to data accuracy, and developed a scheme to assign reliability score to each of them: high (=3), acceptable (=2), and unacceptable (=1). This way, it is possible to determine the relative degree of reliability of landscape indices (Table 1). In general, more factors need to be considered for comparing different landscapes at one point of time or quantifying changes of a landscape at different times than for analyzing one landscape at one point of time. Here we listed three primary factors for one-time landscape analysis: map quality, error balance, and spatial resolution. Other factors such as

Table 1 An example: empirical determination of the relative reliability of landscape indices derived from remotely sensed data

Considered factor		Reliability score		
		3	2	1
One-time landscape analysis	Map quality	$OA > 90\%$	80–90%	$OA < 80\%$
	Error balance	$\lvert UA - PA \rvert_{max} < 5$ for all types	In between	$\lvert UA - PA \rvert_{max} > 15$ for all types
	Spatial resolution	$MMU > 2PS$	In between	$MMU < PS$
Landscape change detection or between-landscape comparison	Consistency in classification accuracy	$\lvert OA_1 - OA_2 \rvert < 5$	In between	$\lvert OA_1 - OA_2 \rvert > 15$
	Consistency in spatial resolution	$PS_1 = PS_2$	In between	$\lvert PS_1 - PS_2 \rvert > (PS_1 + PS_2)/3$
	Consistency in data processing	Same procedures used for all data processing	Different procedures used for some data processing	Different procedures used for each data processing
	Mean accuracy ($\overline{OA}$)	$\lvert PC \rvert > 2(1 - \overline{OA})$	In between	$\lvert PC \rvert < (1 - \overline{OA})/2$
	Consistency in seasonality	All data from the same season	Some data from different seasons	All data from different seasons (e.g., leaf on vs. leaf off)
	Consistency in MMU	$MMU_1 = MMU_2$	$MMU_1 \cong MMU_2$	$MMU_1 \neq MMU_2$

OA = Overall classification accuracy (%); UA = Users' classification accuracy (%); PA = Producer's classification accuracy (%); PS = Pixel size of remotely sensed data (m^2); MMU = Minimum mapping unit (m^2); PC = Potential change rate in area for all the land types (%)

vegetation of interest (forest vs. desert or canopy layer vs. understory layer), landscape contrast (continuous vs. abrupt changes), and patch type characteristics (within-patch variability) may all be important, but are hard to quantify. For between-landscape comparison or landscape change detection, at least six additional factors need to be considered: consistency in classification accuracy, consistency in spatial resolution, consistency in data processing, mean accuracy $\left(\overline{OA}\right)$, consistency in seasonality, and consistency in *MMU*. The overall reliability score (RS) can be obtained by summing up the reliability scores of all factors: $RS = \sum_{i=1}^{n} S_i$, where n is the number of factors considered and S_i is a reliability score found in Table 1.

Concluding remarks

No map is without errors. As landscape ecologists, we need to know where the errors occur and how large they are. The use of remote sensing data will certainly continue to grow, and so will the use of landscape metrics for characterizing spatial patterns. Computerized classification algorithms or classifiers deal with mainly the "color" of the image data, and are still limited in the use of textural (spatial) information. From this point of view, digital classification is far less intelligent than human interpretation that can flexibly integrates various kinds of information. From the selection of image sources to the interpretation of landscape metrics, professional knowledge of the landscape under study is always critical, and cannot and should never be substituted by machine-driven macros.

The problem of classification accuracy and its influence on landscape metrics must be dealt with more explicitly and more rigorously to make landscape pattern analysis ecologically more relevant and effective (Li and Wu 2004, 2007). Unfortunately, except for the few important studies in the 1990s (e.g., Hess 1994; Hess and Bay 1997; Wickham et al. 1997), this critical problem has largely been ignored in landscape ecology. As the issues of data acquisition and accuracy assessment are among the key topics in landscape ecology (Wu and Hobbs 2002; Iverson 2007), we hope that this paper will help stimulate more research interests to elevate our understanding of these issues to a new level.

Acknowledgments We thank four anonymous reviewers for their critical comments on an earlier version of the paper. JW's research was supported in part by the National Science Foundation under Grant No. BCS-0508002 (Biocomplexity/CNH) and under Grant No. DEB-0423704, Central Arizona-Phoenix Long-Term Ecological Research (CAP LTER).

Reference

Anderson JR, Hardy EE, Toach JT, Witmer RE (1976) A land use and land cover classification system for use with remote sensor data, U.S. Geological Survey Professional Paper 964, U.S. Government Printing Office, Washington, DC 28 p

Arnot C, Fisher PF, Wadsworth R, Wellens J (2004) Landscape metrics with ecotones: pattern under uncertainty. Landsc Ecol 19:181–195

Benz UC, Hofmann P, Willhauck G, Lingenfelder I, Heynen M (2004) Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J Photogram Rem Sens 58:239–258

Congalton RG, Green K (1999) Assessing the accuracy of remotely sensed data: principles and practices. Lewis Publishers, New York

Fang S, Gertner G, Wang G, Anderson A (2006) The impact of misclassification in land use maps in the prediction of landscape dynamics. Landsc Ecol 21:233–242

Groom G, Mucher CA, Ihse M, Wrbka T (2006) Remote sensing in landscape ecology: experiences and perspectives in a European context. Landsc Ecol 21: 391–408

Hansen MC, Reed B (2000) A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products. Int J Rem Sens 21:1365–1373

Hess GR (1994) Pattern and error in landscape ecology: a commentary. Landsc Ecol 9:3–5

Hess GR, Bay JM (1997) Generating confidence intervals for composition-based landscape indexes. Landsc Ecol 12:309–320

Iverson L (2007) Adequate data of known accuracy are critical to advancing the field of landscape ecology. In: Wu J, Hobbs R (eds) Key topics in landscape ecology. Cambridge University Press, Cambridge, UK, pp 11–38

Jelinski DE, Wu J (1996) The modifiable areal unit problem and implications for landscape ecology. Landsc Ecol 11:129–140

Kettig RL, Landgrebe DA (1976) Classification of multispectral data by extraction and classification of homogeneous objects. IEEE Trans Geosci Electron GE-14:19–26

Lathrop RG, Montesano P, Haag S (2006) A multi-scale segmentation approach to mapping seagrass habitats using airborne digital camera imagery. Photogram Eng Rem Sens 72:665–675

Li H, Wu J (2004) Use and misuse of landscape indices. Landsc Ecol 19:389–399

Li H, Wu J (2007) Landscape pattern analysis: key issues and challenges. In: Wu J, Hobbs R (eds) Key topics in landscape ecology. Cambridge University Press, Cambridge, UK, pp 39–61

Lu DS, Mausel P, Batistella M, Moran E (2004) Comparison of land-cover classification methods in the Brazilian Amazon Basin. Photogram Eng Rem Sens 70:723–731

Moody A, Woodcock CE (1995) The influence of scale and the spatial characteristics of landscapes on land-cover mapping using remote sensing. Landsc Ecol 10:363–379

O'Neill RV, Krummel JR, Gardner RH, Sugihara G, Jackson B, DeAngelis DL, Milne BT, Turner MG, Zygnut B, Christensen SW, Dale VH, Graham RL (1988) Indices of landscape pattern. Landsc Ecol 1:152–162

Powell RL, Matzke N, de Souza C, Clark M, Numata I, Hess LL, Roberts DA, Clark M, Numata I, Hess LL, Roberts DA (2004) Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. Rem Sens Env 90:221–234

Saura S (2004) Effects of remote sensor spatial resolution and data aggregation on selected fragmentation indices. Landsc Ecol 19:197–209

Scepan J (1999) Thematic validation of high-resolution global land-cover data sets. Photogram Eng Rem Sens 65: 1051–1060

Shao G, Wu W (2004) The effects of classification accuracy on landscape indices. In: Lunetta RS, Lyon JG (eds) Remote sensing and GIS accuracy assessment. CRC Press, Boca Raton, FL, pp 209–220

Shao G, Liu D, Zhao G (2001) Relationships of image classification accuracy and variation of landscape statistics. Can J Rem Sens 27:33–43

Shao G, Wu W, Wu G, Zhou X, Wu J (2003) An explicit index for assessing the accuracy of cover class areas. Photogram Eng Rem Sens 69:907–913

Shen W, Jenerette GD, Wu J, Gardner RH (2004) Evaluating empirical scaling relations of pattern metrics with simulated landscapes. Ecography 27:459–469

Turner MG (2005) Landscape ecology: what is the state of the science? Ann Rev Ecol Evol Syst 36:319–344

Turner MG, O'Neill RV, Gardner RH, Milne BT (1989) Effects of changing spatial scale on the analysis of landscape pattern. Landsc Ecol 3:153–162

Vogelmann JE, Stephen M, Howard M, Yang L, Larson CR, Wylie BK, van Briel 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. Photogram Eng Rem Sens 67:650–662

Wickham JD, Riitters KH (1995) Sensitivity of landscape metrics to pixel size. Int J Rem Sens 16:3585–3594

Wickham JD, O'Neill RV, Riitters KH, Wade TG, Jones KB (1997) Sensitivity of selected landscape pattern metrics to land-cover misclassification and differences in land-cover composition. Photogram Eng Rem Sens 63:397–402

Wu J (2004) Effects of changing scale on landscape pattern analysis: scaling relations. Landsc Ecol 19:125–138

Wu J, Hobbs R (2002) Key issues and research priorities in landscape ecology. Landsc Ecol 17:355–365

Wu J, Jones KB, Li H, Loucks OL (eds) (2006) Scaling and uncertainty analysis in ecology: methods and applications. Springer, Dordrecht, The Netherlands

Wu W, Shao G (2002) Optimal combinations of data, classifiers, and sampling methods for accurate characterizations of deforestation. Can J Rem Sens 28:601–609