

## Integration of forest inventory and satellite imagery: a Canadian status assessment and research issues

Tarmo K. Remmel<sup>a,\*</sup>, Ferenc Csillag<sup>a</sup>, Scott Mitchell<sup>b</sup>, Michael A. Wulder<sup>c</sup>

<sup>a</sup>*Department of Geography, University of Toronto at Mississauga, 3359 Mississauga Road, Mississauga, Ont., Canada L5L 1C6*

<sup>b</sup>*Department of Geography and Environmental Studies, Carleton University, B349 Loeb Building,  
1125 Colonel By Drive, Ottawa, Ont., Canada K1S 5B6*

<sup>c</sup>*Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, 506 Burnside Road, Victoria, BC, Canada V8Z 1M5*

Received 12 August 2004; received in revised form 15 November 2004; accepted 15 November 2004

### Abstract

Canada's ability to sustainably manage approximately 10% of the global forest cover is a critical environmental and economic issue. The capacity to meet such demands and to deliver on national and international commitments regarding forest management is enabled through collaboration between federal, provincial, and territorial agencies. A principal collaborator is the National Forest Inventory (NFI); a systematic photo-plot based monitoring system designed specifically for reporting purposes and as an important input for scientific models. Satellite imagery is illustrated here as a support data set to ensure the quality of the NFI, for auditing the photo-plot contents, and to detect spatial biases. The Canadian Forest Service, in collaboration with the Canadian Space Agency and other federal and provincial agencies, is producing a national land cover database of the forested area of Canada (Earth Observation for Sustainable Development of Forests (EOSD)) using Landsat-7 ETM+ data for circa 2000 conditions. The integration between the plot-based NFI with classified EOSD data is presented for central British Columbia, an area comprising 6 Landsat scenes and 324 2 km × 2 km photo-plots. Traditional accuracy assessment measures based on the analysis of coincidence matrices are reported as levels of agreement for hierarchically aggregated land cover categories (overall agreements of 91%, 79%, 64% and 26% for 3, 4, 6 and 20 classes respectively) to demonstrate coincidence between the different data products. Local agreement between NFI and EOSD is demonstrated as a means of photo-plot auditing while spatial biases are detected through investigations of geographic pattern in the coincidence values. The illustrated approaches may be expanded or applied to different mapped attributes (e.g., biomass) that are of utility to those attempting to characterize large areas in a consistent and rigorous fashion.

© 2004 Elsevier B.V. All rights reserved.

**Keywords:** Coincidence matrix; Agreement; NFI; EOSD; Aggregation; Mutual information; Reliability; Uncertainty coefficient

### 1. Introduction

Canadian forests, representing approximately 10% of the world's forest inventory, play a significant role in the national economy, pose challenges for

\* Corresponding author. Tel.: +1 905 828 3868;  
fax: +1 905 828 5273.

E-mail address: [tarmo.remmel@utoronto.ca](mailto:tarmo.remmel@utoronto.ca) (T.K. Remmel).

environmental management, and are the focus of discussions regarding meeting national and international commitments related to carbon cycling (Natural Resources Canada, 2003). The vast distribution, utilization, and protection of forests across this expansive country dictate a need for continual and timely forest inventory updates to track this dynamic supply of renewable resources (Wulder et al., 2004b). Forest inventory methods in Canada differ by political jurisdiction, intended purpose, and forest type (Gillis and Leckie, 1993) leading to fragmented and often incompatible data products that do not easily aggregate nationally.

The National Forest Inventory (NFI) has been developed to enable a standardized approach for the attribution and collection of forest characteristics nationally on a regular basis. The NFI enables standardization of data collection protocols, collected attributes, and reporting abilities and is the main mechanism through which national statistics and reports are compiled to represent Canada's forests. Further, the NFI provides otherwise unavailable nationally consistent data in support of programs (e.g., carbon budget modeling) requiring information regarding forests (Wulder et al., 2004b). Although it is important to ensure that the NFI is accurate, free of bias, and that it meets prescribed collection standards, we emphasize that the NFI should also integrate effectively with newly developed data products. Comparison of the NFI with classified satellite imagery enables meeting some of these needs because the imagery provides independent measures of characteristics (Hyypä et al., 2000) that may be compared with the NFI photo-plot attributes.

The goal of this research is to extend traditional accuracy assessment techniques, applied as map comparison tools, to the integration of satellite imagery with forest inventory data. The objectives are to illustrate the potential of two-directional analyses and the complimentary nature of two national forest survey data products. Furthermore, identification of geographic variability in the agreement between the data products might indicate regions of varying integration potential where additional sampling efforts can improve integration. To successfully integrate the NFI and the satellite derived Earth Observation for Sustainable Development of Forests (EOSD) land cover products we first assess their

similarities and differences (agreement). We begin with a traditional coincidence matrix approach based on accuracy assessment tools developed for the remote sensing community, as these are currently the best and most convenient techniques for the comparison of categorical data sets. Thus, implementation of accuracy assessment protocols between the data sets provides global surrogate measures of agreement. Although the accuracy assessment protocols are a necessary step towards integration, they are insufficient to fully characterize the thematic and spatial controls on this agreement. Our emphasis is to develop an intelligent approach for integrating information from both the NFI and the EOSD, such that each might contribute to the overall knowledge gain rather than assuming that one data set is more accurate than the other. This creates a two-way exchange of information, exploring questions such as “does knowing the EOSD provide valuable information about the NFI?” and vice versa. Such a comparison will provide guidance for an actual field-based accuracy assessment. We explore subsequent techniques for characterizing agreement specifics by unraveling some of the spatial, geographic, and thematic controls with the intent of identifying reasons for disagreements and for locating spatial biases that may be apparent in the NFI.

We conduct the global assessment of agreement and further account for the individual and combined impacts of thematic detail, geographical variation in the distribution of classes, and spatial heterogeneities in landscape pattern. The following section provides background material to enable an understanding of the national land cover mapping programs under consideration. These are followed by a brief discussion of accuracy assessment protocols for the agreement between the raster (EOSD) and vector (NFI) data sets.

## *1.1. Background*

### *1.1.1. Study area*

This case study focuses on a region centered on the city of Prince George (53°53'N, 122°40'W) located in central British Columbia, Canada. This region represents approximately 14% of the province by area and exhibits varying topographic and land cover characteristics. This region has a characteristic east-west gradient in topography and weather, with the

western boundary exhibiting more significant and variable relief than the alpine plateaus to the east. The implications are wetter conditions to the west, potentially influencing the land cover composition and distribution within an area characterized by a continental climate and moderate precipitation.

Forest management and to a lesser extent agriculture are the primary anthropogenic activities affecting the landscape, while fire and insect infestations yield much of the natural disturbance within the region. This area encompasses several large communities, sections of mountain ranges, and sizable plateaus (Meidinger and Pojar, 1991). Forest species dominant within the Prince George area are hybrid white spruce (*Picea glauca* × *engelmannii*), sub-alpine fir (*Abies lasiocarpa*), lodgepole pine (*Pinus contorta*), and Douglas-fir (*Pseudotsuga menziesii*) with minor constituents of black spruce (*Picea mariana*), trembling aspen (*Populus tremuloides*), balsam poplar (*Populus balsamifera*), and paper birch (*Betula papyrifera*) (Meidinger and Pojar, 1991).

#### 1.1.2. National Forest Inventory (NFI)

A consortium of independent provincial agencies conducts the compilation of the NFI from forest inventories. Collectively, the NFI aims to monitor and assess the extent and state of Canadian forests in a timely and accurate manner. Individual provinces and territories in Canada have stewardship responsibilities for natural resources. Each individual jurisdiction has control over the definition and collection of forest attributes. Consequently, the techniques and attributes collected are not necessarily consistent across the country. The CFS has drafted a framework for compiling 25 desired attributes that serve as a minimum reporting level for the individual provinces and territories (Gillis, 2001). These attributes, generally related to physical, ground level measurements of forest type, species composition, volume, biomass, and disturbance are common parameters used by foresters but often difficult to quantify with remotely sensed data. Both vegetated and non-vegetated land cover polygons within the NFI database are coded hierarchically, such that attributes can be aggregated, providing multiple levels of detail for mapping. The NFI is based on a 6-level hierarchical classification (Table 1), of which five levels are reflected in the EOSD (Wulder and Nelson, 2003), that accounts for

vegetation cover, tree cover, landscape position, vegetation type, forest density, and species diversity (Natural Resources Canada, 1999; Gillis, 2001). The corresponding EOSD land cover legend reflects the hierarchical structuring of the NFI classification scheme allowing the integration of the EOSD and NFI data (Wulder and Nelson, 2003).

The NFI sampling structure is based on the landmass of Canada being overlain with a regular grid with a spacing of 20 km (based on the Albers equal-area projection). At the nodes of this national grid, for forested areas, are 2 km × 2 km photo-plots, forming an approximate 1% national sample for representing Canada's forests. At each photo-plot location, detailed forest attributes have been compiled and stored within a digital vector GIS database (Gillis, 2001). Nationally there are approximately 20,000 of these photo-plots, 324 of which reside within our study area. These 324 photo-plots from the year 2000 are comprised of 11,049 aerial photo interpreted polygons to which the attributes are assigned. Details of the NFI plot-based design are available (Natural Resources Canada, 1999; Gillis, 2001; Wulder et al., 2001), which outline the requirements for sub-sampling, and locating photo-plots and ground plots. The NFI sampling structure enables a mechanism for harmonization of forest monitoring for national reporting in Canada.

#### 1.1.3. Earth observation for sustainable development of forests (EOSD)

The Canadian Forest Service (CFS) and the Canadian Space Agency are undertaking a joint project with provincial, territorial, and federal agencies to map the forested area of Canada with satellite imagery. The project, entitled Earth Observation for Sustainable Development of Forests (EOSD) is producing a land cover classification representing circa year 2000 conditions of the land cover of the forested area of Canada for completion in early 2006 (Wulder et al., 2003). Experimental EOSD mapping focuses primarily on forested areas (Wulder and Seemann, 2001), based on images consisting of 30 m spatial resolution orthorectified Landsat-7 ETM+ data produced for the entire Canadian landmass (Wulder et al., 2002).

Six Landsat ETM+ scenes (P = Path, R = Row), P48/R22, P48/R23, P49/R22, P49/R23, P50/R22 and P50/R23 form the study area for which a 22-category EOSD classification is shown in Fig. 1. The images

Table 1

Tabular depiction of the relationship between EOSD labels, hierarchically structured NFI attributes, and the four aggregation levels used in this study to assess the impact of thematic detail on accuracy assessment

Aggregate Classes					Linking Categories		NFI Attributes												
AGG3		AGG4		AGG6		AGG20		EOSD		Density	VegType	Position	Cover	Base					
						1	Shadow												
						2	Cloud												
1	Water	1	Water	1	Water	6	Water	6	Water				Water	Non-vegetated					
2	Land	2	Land	2	Land	3	Snow/Ice	3	Snow/Ice		Snow/Ice	Wetland Upland Alpine	Land						
						4	Rock/Rubble	4	Rock/Rubble		Rock								
						5	Exposed Land	5	Exposed Land		Exposed								
3	Vegetated	3	Non-forest Vegetated	3	Non-forest Vegetated	7	Shrub-Tall	7	Shrub-Tall		Tall Shrub	Upland Alpine	Non-treed	Vegetated					
						8	Shrub-Low	8	Shrub-Low		Low Shrub								
						9	Herb	9	Herb		Herb								
						10	Bryoids	10	Bryoids		Bryoid								
						11	Wetland-Treed	11	Wetland-Treed			Wetland	Treed						
						12	Wetland-Shrub	12	Wetland-Shrub		Shrubs	Wetland	Non-treed						
						13	Wetland-Herb	13	Wetland-Herb						Herb				
						4	Forest	4	Conifer		14	Conifer-Dense	14		Conifer-Dense	Dense	Conifer	Upland	Treed
											15	Conifer-Open	15		Conifer-Open	Open			
		16	Conifer-Sparse	16	Conifer-Sparse					Sparse									
		5	Broadleaf	17	Broadleaf-Dense			17	Broadleaf-Dense	Dense	Broadleaf								
				18	Broadleaf-Open			18	Broadleaf-Open	Open									
		6	Mixed	6	Mixed	19	Broadleaf-Sparse	19	Broadleaf-Sparse	Sparse	Mixed								
						20	Mixed-Dense	20	Mixed-Dense	Dense									
21	Mixed-Open					21	Mixed-Open	Open											
				22	Mixed-Sparse	22	Mixed-Sparse	Sparse											

EOSD categories 1 and 2 do not correspond with the NFI; thus, they are not included in the aggregate classes. Thematic detail increases towards the center of the table as approached from the outside edges.

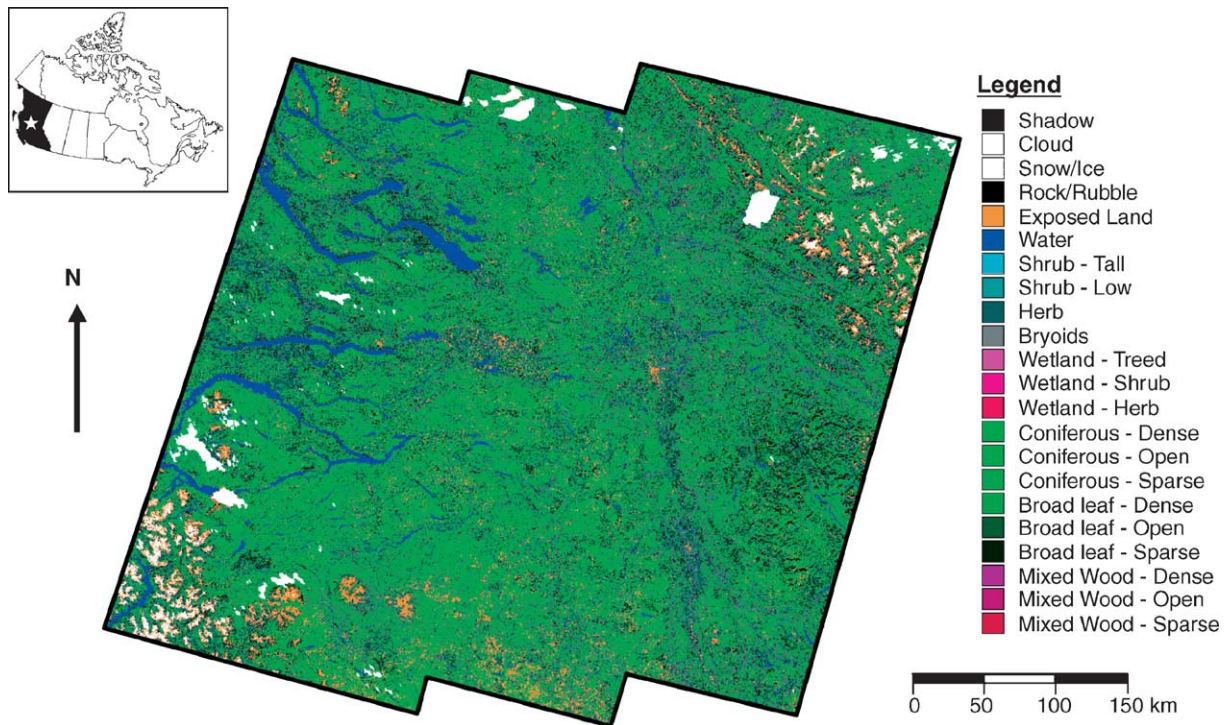


Fig. 1. The 22-category EOSD forested land cover classification data for the study area centered on Prince George, BC, Canada. The star on the inset map indicates the approximate location of the study area.

were acquired on 23 September 2000, 12 September 1999, and 2 August 1999 for Landsat Paths 48, 49, and 50, respectively. An unsupervised clustering algorithm forms the core of the classification procedure used to produce the EOSD. A combination of an NDVI masking layer, a  $3 \times 3$  variance-texture layer computed from the Landsat-7 panchromatic channel, and the six optical ETM+ layers are submitted to a  $k$ -means classifier that generates nearly 250 classes, which are subsequently merged into user dictated land cover classes (Wulder et al., 2004a, 2004b). The classes are assigned labels according to the hierarchical legend defined to link with the NFI classification (Wulder and Nelson, 2003) using the best available data, including forest inventory data, air photos, plot data, and local knowledge (Wulder et al., 2003).

#### 1.1.4. Accuracy assessment protocols for determining agreement

The core purpose of accuracy assessment is to describe quantitatively the coincidence between two independently collected data sets (Foody, 2002),

generally beginning with the construction of a coincidence matrix (also referred to as an error, confusion, or contingency matrix) (Congalton, 1991; Stehman, 1997b). Due to the diversity of questions that can be asked about accuracy, there is no widely accepted single standard measure for its characterization (Stehman and Czaplewski, 1998; Foody, 2002), and accuracy reporting for large-area land cover map products is particularly challenging. The sophistication in accuracy reporting is emphasized by the recognition of different accuracy components. The statistically most rigorous developments are related to deriving confidence intervals for popular measures computed from the coincidence matrix, which are largely based on stochastic models known from survey statistics (Stehman, 1997b; Stehman and Czaplewski, 2003). Over the last decade, emphasis has been focused on variations in, and controls of, coincidence (or lack thereof). Foody (2002) summarizes most of these developments that range across the widely debated kappa coefficient of agreement (Monserud and Leemans, 1992), applying fuzzy class member-



ships (Gopal and Woodcock, 1994), accounting for geographical trends and local variation (Kyriakidis and Dungan, 2001), thematic level of detail (Yang et al., 2001; Rogan et al., 2003), and separating location from quantification errors (Pontius, 2000). More recently, accuracy assessment finds itself as a subset of categorical map comparison (Patil et al., 2004; Csillag and Boots, 2004).

The realm beyond simple coincidence matrix evaluations provides an opportunity to make specific statements about the agreement between data sets. These statements carry much more information than simple match/mismatch proportions that are typically reported as the overall agreement. Consideration of off-diagonal elements with an entropy-based approach provides insight to the bi-directional influence of individual categories to the agreement between data sets (including rare categories). Observation of geographic variability in these and traditional measures provide insight into possible spatial and categorical biases in the agreement/disagreement.

## 2. Methods

### 2.1. Linking NFI and EOSD

The NFI and EOSD are two Canadian wide-area land cover products, each compiled independently using different methods and resulting in different (but purposefully related) thematic attributes. Integration of these two data sets requires alignment of their attributes making them comparable within a common classification regime. We provide only a brief overview of the classification category descriptions here and refer the reader to existing and extensive documentation on this subject (Gillis, 2001; Wulder and Nelson, 2003). However, we discuss the linkage between the two systems in detail.

The most detailed categories in the EOSD land cover legend are slightly cruder than their NFI counterparts; thus, the EOSD categories form the standard for linking the two data products. To achieve this objective, the polygonal NFI inventory needed assignment of a new attribute corresponding to the most representative EOSD category. Furthermore, conversion of polygons to raster facilitated direct comparisons with the experimental EOSD data

provided at 30 m spatial resolution. The terminology of this paper adheres to the following conventions when referring to these data products and attributes: (1) the original EOSD data classification labels will be termed 'EOSD', and (2) the EOSD label attached to the NFI data will be called '*NFI*'. Thus, when discussing the data product in general, the acronym is not italicized, but to specify discussion regarding the actual label assigned to the data product, we use the italicized form.

The EOSD land cover legend comprises 23 categories; however, category 0 represents no data, and categories 1 and 2 are shadow and cloud, respectively. Since these three categories are not present in the NFI, they are dropped from the EOSD–NFI alignment process. The remaining categories, related to vegetated and non-vegetated plots, match based on a series of hierarchical relationships (Table 1).

Software written to systematically assign the proper *NFI* labels to the existing polygons follows a classification-tree construct. Polygon sets were selected that met the criteria for a complete branch of the classification tree; then the corresponding label was assigned to those polygons before selecting the next set of polygons analogous to the next branch of the classification tree. This routine was repeated until all possible EOSD labels had been assigned. Polygon selection first considered the non-vegetated locations, followed by those that were vegetated (forming the first split in the hierarchical classification). Non-vegetated sites were further scrutinized based on finer divisions of water or land categories, while vegetated sites were split based on being treed or non-treed. These classes were subsequently separated into wetland, upland, or alpine categories. Further scrutiny related to vegetation types (conifer, broadleaf, mixed) and density classes (dense, open, sparse) formed the finest levels of the classification tree.

The envisioned integration of EOSD and NFI could utilize detailed information about the coincidence between the two data sets when moving along the classification hierarchy (e.g., from a few to many classes). Therefore, we have defined four thematic aggregations of land cover classification for the assessment of agreement. These aggregation levels are termed AGG3, AGG4, AGG6, and AGG20 to reflect the aggregation to a specified number of

thematic categories (Table 1). The EOSD (linking categories) column in Table 1 represents the full compliment of categories. The remaining columns to the left refer to the aggregate category labels and their corresponding numeric identifiers and the columns to the right indicate the hierarchical coding of the NFI attributes. Thus, the independent thematic labels of the NFI and EOSD can be compared via this alignment, resulting in a common classification legend.

Although image rectification and registration are ongoing issues with studies involving remotely sensed data, the spatial biases that are of interest to the current monitoring program are not image processing and rectification based, but related to the actual manual interpretation and attribution of the photo-plots. Different interpreters may have class or category preferences that are captured in the photo-plots and if these preferences were distributed randomly, their impacts may not be evident. The interpretation process typically employs a single interpreter or agency to interpret a block of land, which can introduce spatial biases in categorization if indeed there are any interoperation preferences, but spatial mis-registration, which also may vary geographically, are of limited effect in this case.

## 2.2. Analysis of coincidence matrices

### 2.2.1. Global agreement

We follow the general approach to accuracy assessment as a surrogate for comparing classified remotely sensed maps, considering response design, sampling design, analysis, and estimation (Stehman, 1996; Stehman and Czaplewski, 1998; Zhu et al., 2000; Foody, 2002). The primary focus in our assessment is on the agreement between EOSD and NFI at the individual pixels because they are the ‘mapping units’ of the national EOSD coverage. Given the varied spectrum of opinions (Rosenfield et al., 1982; Aronoff, 1982; Congalton, 1991; Edwards et al., 1998) about the level of statistical rigor in the construction and analysis of the coincidence matrix, we briefly review the basic principles and describe the methodology.

The response and sampling designs for this study are based on the NFI. The first-stage spatial (cluster) sampling design results from the regularly spaced photo-plots and the hierarchical classification system

Table 2

Sample  $2 \times 2$  contingency table and notation for the analysis of coincidence between NFI and EOSD

	Coincide	Mismatch	Total
NFI + EOSD	$k$	$M - k$	$M$
EOSD – NFI	$n - k$	$N - M - n + k$	$N - M$
Total	$n$	$N - n$	$N$

NFI + EOSD represents all locations where both data sets are available (“the sample”) and EOSD–NFI corresponds to all locations where the reference data are not available. The analysis aims to estimate  $n$  (or  $n/N$ ) while knowing  $k$  (or  $k/M$ ).

(summarized in Table 1 and Fig. 1), which lead to an extensive (but not exhaustive) polygon coverage with labels corresponding to those of the EOSD. These polygons are subsequently converted into a grid conforming to the EOSD with 30 m spatial resolution. The overall design principles aim at ensuring that the NFI photo-plots represent the entire landscape. These principles manifest themselves in assumptions that may or may not be easy to test and which have implications for the statistical analysis of agreement.

The statistical evaluation of the match between EOSD and NFI (“the agreement”) is based on the overall coincidence matrix with appropriately created contingency tables representing coincidence and mismatch between the two data sets (Table 2). These tables are simple and straightforward to analyze when generated by simple random sampling (SRS). The basic model of SRS (i.e., that each pixel has equal chance to get into the sample – also referred to as the “urn model”) is the simplest to implement computationally,<sup>1</sup> but the most challenging to execute in reality. However, it is important to point out that the SRS assumptions are only required for estimating confidence intervals appropriately (i.e., estimating the proportion of coincidence, “the agreement”, does not require it). Let  $N$  denote the total area (in our case the Prince George study area of 146,417,077 pixels) and let  $M$  denote the sample size (an approximate 1% sample in our case equaling 1,438,782 pixels) where both NFI and EOSD category labels are known and coincide in  $k$  cases (Table 3). When  $N$  and  $M$  are fixed

<sup>1</sup> All computations were conducted using R (CRAN, 2004). The R project for statistical computing. <http://www.r-project.org>; Bivand, R.S., 2000. Using the R statistical data analysis language on GRASS 5.0 GIS database files. Comput. Geosci., 26, 1043–1052).

Table 3

Thematic distributions of NFI and EOSD data products in terms of area and percentage cover

Category	NFI		EOSD		NFI		EOSD	
	Total polygon area		Total grid area		Grid area within NFI		Grid area within NFI	
	ha	%	ha	%	ha	%	ha	%
1 Shadow	0.0	0.0	397330.0	3.0	0.0	0.0	4139.7	3.3
2 Cloud	0.0	0.0	47.9	0.0	0.0	0.0	0.1	0.0
3 Snow/ice	0.0	0.0	133924.6	1.0	0.0	0.0	1190.1	0.9
4 Rock/rubble	7624.5	5.9	0.0	0.0	7616.1	5.9	0.0	0.0
5 Exposed land	1290.9	1.0	948731.9	7.2	1289.8	1.0	8565.0	6.7
6 Water	4961.4	3.8	532824.4	4.0	4956.4	3.8	4376.1	3.4
7 Shrub – tall	3409.8	2.6	80534.9	0.6	3406.1	2.6	906.7	0.7
8 Shrub – low	7789.6	6.0	755255.3	5.7	7788.0	6.0	7483.0	5.9
9 Herb	2058.7	1.6	708614.4	5.4	2059.9	1.6	6655.6	5.2
10 Bryoids	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11 Wetland – treed	1612.5	1.2	0.0	0.0	1609.3	1.2	0.0	0.0
12 Wetland – shrub	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13 Wetland – herb	3006.8	2.3	2183.5	0.0	2992.1	2.3	20.2	0.0
14 Conifer – dense	10539.8	8.1	3728787.6	28.3	10537.7	8.1	37827.5	29.7
15 Conifer – open	66063.4	51.0	3331776.0	25.3	66066.6	51.0	31116.9	24.5
16 Conifer – sparse	8973.8	6.9	331609.3	2.5	8964.2	6.9	3012.7	2.4
17 Broadleaf – dense	1107.4	0.9	439567.6	3.3	1106.0	0.9	4712.9	3.7
18 Broadleaf – open	4093.2	3.2	191338.3	1.5	4083.8	3.2	1930.7	1.5
19 Broadleaf – sparse	153.3	0.1	800033.7	6.1	154.7	0.1	7530.1	5.9
20 Mixed – dense	680.0	0.5	690293.0	5.2	678.9	0.5	6724.9	5.3
21 Mixed – open	4394.6	3.4	104684.9	0.8	4393.9	3.4	1039.1	0.8
22 Mixed – sparse	1786.7	1.4	0.0	0.0	1787.1	1.4	0.0	0.0
Total	129546.4	100.0	13177536.9	100.0	129490.4	100.0	127231.0	100.0

This table provides the breakdown by class membership of the discrete photo-plot-derived NFI and the continuous satellite-derived EOSD. Furthermore, the raster area, by thematic class, of the NFI and EOSD are provided for the area constrained to be within the NFI photo-plots.

and  $n$  is known, the probability distribution function (PDF) of  $k$  is hypergeometric over  $k = 0, 1, \dots, n$ , which can be approximated by a binomial PDF of order  $M$  and  $p = n/N$ . However, in our case,  $k$  is known and  $n$  (consequently  $p$  with the binomial approximation) is unknown. The PDF of  $p$  is a beta-distribution (with parameters  $M - k + 1$  and  $k + 1$ ) for which confidence intervals can be computed. It is important to note that the confidence intervals depend on the sample size ( $M$ ) and not the proportion of the sample ( $M/N$ ). Higher sampling proportions usually comply better with the requirements of equal probability sampling (SRS) and in large-area surveys, lead to extremely tight confidence intervals (Stehman, 2001).

A  $2 \times 2$  contingency table can be constructed for any collection, combination, or aggregation of categories. When all categories are considered, it is usually referred to as an “overall agreement”, which represents the “average” (area-weighted) proportion of coincidence. Individual categories can also be

considered in the same manner, in which case they can be evaluated from a user’s perspective (relative to the distribution of EOSD categories) or the producer’s perspective (relative to the distribution of NFI categories). For the overall and user agreements, more efficient estimates can be obtained by post-stratification, by which coincidence proportions in the sample are re-weighted according to their observed relative frequencies in the total population (Card, 1982; Zhu et al., 2000). The (regular) grid of the reference photo-plots corresponds to a one-stage geographically stratified cluster sampling design (Stehman, 1997a; Zhu et al., 2000), but we apply the SRS formulae for the following reasons. First, the sampling design does not violate the equal-probability property. Second, since ignoring the clusters will underestimate the variance (of  $n/N$ ) and ignoring the geographical stratification will overestimate it (Zhu et al., 2000), the two are likely to reduce bias by “compensation”. Finally, using the binomial approx-



imation for cluster sampling (Stehman, 1997a) can underestimate the variance (of  $n/N$ ) by two orders of magnitude without explicitly accounting for spatial autocorrelation (Kingston, 2004).

### 2.2.2. Geographic variability in agreement

The agreement measures derived from the coincidence matrix provide global summaries. For national (i.e., large-area) monitoring, it is particularly important to decompose these global measures along various dimensions. Here, we are primarily concerned with the spatial and thematic homogeneity/heterogeneity and their interactions as a corollary of the “cartographic uncertainty relationship” (Csillag, 1991). Global measures, by necessity, are based on the assumption of spatial stationarity, that is, the stochastic characteristics (e.g., composition, configuration) of the observed data, and consequently the error distributions, are assumed constant. It is impossible to prove the validity of this assumption, but we describe here some exploratory tools that can be used as indicators of the degree of deviations.

The NFI sampling design provides the capacity to explore and decompose the spatial distribution of accuracy values. The global measures, as the spatially coarsest extreme, represent overall characteristics for approximately 130,000 km<sup>2</sup>. We can decompose this global accuracy measure by geographical regions, e.g., conveniently into 324 values (one for each NFI photo-plot) and further into 11,049 individual NFI polygons. Ultimately, we can generate a “reliability” value for each of the 146,417,077 pixels. At each level of the geographic decomposition, we can explore the behavior of the chosen characteristics related to accuracy, which usually is some form of the question “what is the probability of getting the right label at this location?”. Given the hierarchical nature of the classification, it is reasonable to expect a monotonic relationship between thematic detail and coincidence between EOSD and NFI: as we aggregate the classes to coarser and coarser themes, coincidence will increase.

### 2.3. Entropy-based analysis of coincidence matrices

Taking the EOSD as a categorical ( $K$  colour) map of the Prince George study area, at any one location the uncertainty about the colour can be characterized by  $H(K)$ , the (overall) entropy of the colours. The overall

entropy characterizes the uniformity of the distribution for colours and takes its maximum value when all colours are equally likely (however, the value of this maximum changes with the number of colours). When we obtain the NFI as a second map with  $L$  colours, we can evaluate if the uncertainty about the EOSD colour changes at any location by knowing the colour on the corresponding NFI map. Generally, when the two maps are similar, the uncertainty is reduced; however, this is not necessarily the case for all colours. The question: “how much more do we know about the NFI given the EOSD?” is essential for this type of inquiry. The nature of the conditional statement also defines a direction (i.e.,  $H(\text{NFI}|\text{EOSD})$  does not necessarily equal  $H(\text{EOSD}|\text{NFI})$ ). This bi-directionality can potentially yield much more useful information than the typical single-directional query when one data product is considered to be more correct than the other is.

At each location, the probability of the NFI colour is  $P(\text{NFI}|\text{EOSD} = k)$ , that is, the conditional probability distribution given that the EOSD map has colour  $k$  at that location (and it is easily computed from the coincidence matrix by dividing each element with the appropriate marginal). The mutual information,  $I(\text{NFI}, \text{EOSD}) = H(\text{NFI}) - H(\text{NFI}|\text{EOSD})$ , characterizes the agreement between the two maps and it can be assessed for significance with  $(K - 1) \times (L - 1)$  degrees of freedom using  $G^2$ , the likelihood ratio chi-square (Goodman and Kruskal, 1979; Freeman, 1987). The overall information gain (on NFI given EOSD) can be conveniently expressed as a percentage called the uncertainty coefficient:  $U(\text{NFI}|\text{EOSD}) = 100 \times (H(\text{NFI}) - H(\text{NFI}|\text{EOSD}))/H(\text{NFI})$ , that is, the relative decrease in uncertainty. The relationship between  $U$  and the overall agreement is highly dependent on the distribution of the off-diagonal elements. The “binomial” approach lumps these into an overall mismatch category, while the “entropy-based” approach weights them according to how uneven the  $P(\text{NFI}|\text{EOSD} = k)$  distribution is. In fact, the decomposition of  $U$  (or even the mutual information between NFI and EOSD) characterizes the contributions of each colour, and this can be particularly useful for rare categories, or categories that behave differently from the overall trend. For example, it can happen that  $P(\text{NFI}|\text{EOSD} = k)$ , the conditional distribution of category  $k$ , is almost uniform (i.e., this category in the EOSD corresponds

with almost equal probability to each category in the NFI). If most of the other conditional distributions are uneven, this category will have negative  $U$  coefficients (i.e., if we find category  $k$  on the EOSD map, our uncertainty about the category on the NFI map increases).

As with the analysis of coincidence matrices, the entropy-based analyses can also be conducted on the individual 324 NFI photo-plot locations. Since the entire Prince George study area is not homogeneous (stationary), realistically the entropy decompositions will differ among individual NFI photo-plots. We illustrate how to interpret such results.

## 2.4. Data characteristics

### 2.4.1. NFI

It is well known that land cover interfaces can cause spectral mixtures that confuse classifiers leading to disagreements between data sets. We hypothesize that if we remove all edge pixels from the analysis and recomputed the overall agreement measures that we should see an improvement. This was conducted by shrinking all NFI polygons using an internal buffer of 50 m prior to converting them to raster, whereby the edge pixels (and in some cases entire small polygons) were eliminated. The remaining areas represented core land cover patches that should be much more homogeneous than the edges. A previous study found that weighting pixel values by their distance to patch borders was not overly successful (Mäkelä and Pekkarinen, 2001), likely due to small patch sizes. Although severe, our approach eliminates small patches, but can also yield global, photo-plot, and individual polygon level results for larger patches. We illustrate the results for a single photo-plot.

### 2.4.2. EOSD

Data reliability of the satellite derived EOSD is important in controlling the overall agreement between the NFI and EOSD and consequently locating biases. Therefore, characterization of per-pixel label reliability provides insight as to how likely it is that a pixel was assigned a reasonable EOSD label. The conditional probability that a pixel ( $X$ ) is assigned to a specific class ( $C_i$ ) can be approached by computing the distance ( $D$ ) of each pixel to its assigned category's spectral cluster centroid, providing a measure of

labeling reliability according to  $D = P(X|C_i)$ . This distance is measured in multi-dimensional spectral-space according to

$$D = \sqrt{\sum_{i=1}^n \left( \frac{r_i - \bar{m}_i}{s_i} \right)^2}$$

where  $r_i$  is the pixel value in the  $i$ th band,  $m_i$  the cluster average in the  $i$ th band, and  $s_i$  the standard deviation of the cluster in the  $i$ th band. The normalized Euclidean distance of each pixel to its cluster centroid is computed and stored as a new raster layer. The novelty of this approach is in its posterior computation allowing the distances to be computed for each of the 20 labeled classes rather than all original 250 clusters. If a pixel's distance from its centroid is small, the likelihood of belonging to the assigned class is high. Conversely, pixels with large distances to their cluster centroids are more likely to be confused among other classes due to mixing in the multi-dimensional spectral space. Ultimately, the spacing of spectral clusters in multi-dimensional space and their individual variances dictate the expectation for reliable classification. For example, the spectral clusters for very distinct (e.g., water) categories may determine very short distances or due to its distance from other clusters allow for larger distances. Conversely, categories such as wetlands may have more dispersed clusters, allowing for larger spectral distance values. The standardized distances for individual pixels are computed and presented in map form. Their individual values can further be analyzed for significance and be used in locating areas where reliability is low and further attention might be warranted. Although this forms the core of a subsequent paper, a brief example is provided here to illustrate the concept.

Data reliability is also linked with analyses regarding the collection date of satellite imagery. Since consecutive passes of the Landsat sensor in the east-west direction are not temporally continuous, the atmospheric and illumination variability can alter the spectral signatures observed for individual pixels on neighbouring images for the same land cover type, thereby influencing potential disagreements. We have only begun these analyses but would like to indicate their importance and possible control on the observed NFI–EOSD agreements.

Table 4

Coincidence table for AGG3 (pixels), with NFI and EOSD marginals, post-stratification weights (POST), and user/producer accuracies with upper and lower bounds

	NFI			EOSD marginal	POST	%		
	Water	Land	Vegetated			LB	User	UB
EOSD								
1 Water	44349	13123	4374	61846	7408322	73.26	73.64	74.01
2 Land	1986	36953	56228	95167	10541465	40.26	40.56	40.85
3 Vegetated	4361	40396	1165910	1210667	124051980	95.94	95.99	96.03
NFI marginal	50696	90472	1226512	1367680				
%								
UB	87.77	41.17	95.10					
Producer	87.48	40.84	95.06					
LB	87.19	40.52	95.02					

### 3. Results

#### 3.1. Coincidence matrix analysis of agreement

##### 3.1.1. Global agreement analysis

The four coincidence matrices representing the four levels of thematic aggregation are presented in Tables 4–7. Each of these tables presents the coincidence values for all thematic combinations between the EOSD and NFI land cover product pixels, yielding overall agreement estimates of 91.2, 79.4, 64.1, and 26.1% for aggregation levels AGG3, AGG4, AGG6, and AGG20, respectively. These tables also contain the coincidence table marginals, post-stratification weights (Czaplewski and Patterson, 2003),

producer's, and user's agreement values with upper (UB) and lower (LB) bounds indicated, as obtained from a beta distribution (see Section 2.2.1 where rationale is presented). The overall agreement values indicate the total match between the NFI and EOSD labels at a specified level of thematic aggregation and represent “global” (single number) agreement summaries for the Prince George study area. The decrease in agreement with increasing thematic detail (AGG3 → AGG20) is evident.

There are EOSD land cover categories that are not reflected in the NFI; thus, the reference columns for shadow, cloud, and snow/ice are filled with zeros; it is impossible to have coincidence for these categories. Furthermore, bryoids and wetland-shrub categories do

Table 5

Coincidence table for AGG4 (pixels), with NFI and EOSD marginals, post-stratification weights (POST), and user/producer accuracies with upper and lower bounds

	NFI				EOSD marginal	POST	%		
	Water	Land	Non-forest vegetated	Forest			LB	User	UB
EOSD									
1 Water	44349	13123	1170	3204	61846	7408322	73.26	73.64	74.01
2 Land	1986	36953	33792	22436	95167	10541465	40.24	40.53	40.82
3 Non-forest vegetated	2554	22569	61657	80613	167393	17184311	36.18	36.39	36.60
4 Forest	1807	17827	80652	942988	1043274	106867669	90.15	90.22	90.29
NFI marginal	50696	90472	177271	1049241	1367680				
%									
UB	87.77	41.17	35.00	89.93					
Producer	87.48	40.84	34.78	89.87					
LB	87.19	40.52	34.56	89.82					

Table 6

Coincidence table for AGG6 (pixels), with NFI and EOSD marginals, post-stratification weights (POST), and user/producer accuracies with upper and lower bounds

	NFI						EOSD marginal	POST	%		
	Water	Land	Non-forest vegetated	Conifer	Broadleaf	Mixed			LB	User	UB
EOSD											
1 Water	44349	13123	1170	2695	162	347	61846	7408322	73.26	73.63	74.01
2 Land	1986	36953	33792	19385	957	2094	95167	10541465	40.23	40.52	40.81
3 Non-forest vegetated	2554	22569	61657	55088	12919	12606	167393	17184311	36.19	36.40	36.61
4 Conifer	1100	6757	36221	698923	13739	42782	799522	82135254	13.78	13.94	14.11
5 Broadleaf	248	8664	34940	70210	22381	21043	157486	15899328	87.30	87.38	87.47
6 Mixed	459	2406	9491	53198	8429	12283	86266	8833087	13.96	14.18	14.40
NFI marginal	50696	90472	177271	899499	58587	91155	1367680				
%											
UB	87.77	41.17	35.00	38.60	77.79	13.70					
Producer	87.48	40.84	34.78	38.20	77.70	13.47					
LB	87.19	40.52	34.56	37.81	77.62	13.25					

not occur in the study area according to the NFI, thus their reference columns have been left out to simplify the table. Similarly, the bryoids, wetland-treed, and wetland-shrub categories are not mapped by the EOSD classification within the study area, thus their rows were dropped. It is apparent that considerable disagreement is present among density classes and forest types. The majority of forests are classified as conifer and broadleaf, with the mixed class being virtually unrepresented. The exposed land category experiences confusion with numerous categories, but primarily with rock/rubble. The spectral confusion has also resulted in some shrub and wetland sites being classified as exposed land. When the thematic detail of AGG20 is aggregated, the variability of individual classes diminishes and EOSD labels agree much better with those of the NFI. Thus, the EOSD matches well with the NFI for coarse forest classes (e.g., AGG3, AGG4, and to an extent AGG6) across the study area, but expresses significantly more disagreement with detailed density, wetland, and shrub classes. Overall, this indicates that the EOSD is not a simple extension of the NFI, but they remain complimentary yielding that their comparison and integration can provide more information than either data product alone.

### 3.1.2. Local agreement analysis

Subsequent to the global agreement analysis, we spatially decomposed the overall agreement at each

NFI photo-plot, yielding 324 overall agreement estimates for each thematic aggregation level. These results are summarized by a map of bar graphs centered on each NFI photo-plot and represent the spatial distribution of overall agreement within the study area (Fig. 2). Each graph on the map comprises four bars (one for each aggregation level), beginning on the left with AGG3 (the coarsest) and ending on the right with AGG20 (the finest) thematic detail aggregation. The heights of the bars indicate the relative values of spatially partitioned overall agreement. The clear trend of decreasing accuracy with increasing thematic detail is visible and the differences among graphs across the study area indicate the geographic variability in the overall local agreement. Fig. 2 shows very little variability in the overall agreement of AGG3 (black bars) and higher variability in the other thematic aggregation levels. Although the overall agreement drops on each graph from left to right (AGG3 to AGG20), the decrease is not consistent across the study area. Some photo-plots exhibit relatively poor agreement at all thematic aggregations (e.g., SW corner), while others are much better (e.g., western-central region) typified by all bars being tall.

When all categories behave similarly across the study area, the overall agreement translates directly to any thematic class. However, variations in the match between EOSD and NFI for individual categories were

Table 7

Coincidence table for AGG20 (pixels), with NFI and EOSD marginals, post-stratification weights (POST), and user/producer accuracies with upper and lower bounds

	NFI																		EOSD marginal	POST	% LB    User    UB		
	4	5	6	7	8	9	11	13	14	15	16	17	18	19	20	21	22						
EOSD																							
1 Shadow	6832	143	1316	201	244	122	67	329	1810	32567	1454	101	254	19	39	435	64	45997	4414778	0.00	0.00	0.00	
2 Cloud	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	532	0.00	0.00	0.00	
3 Snow/Ice	12922	2	44	28	0	0	0	0	0	187	37	0	1	0	0	2	0	13223	1488051	0.00	0.00	0.00	
5 Exposed land	32610	4343	1986	2106	25960	3488	645	2238	897	12997	5491	190	729	38	47	366	1036	95167	10541465	4.43	4.56	4.70	
6 Water	141	58	44305	195	60	52	159	835	254	2137	80	42	103	16	59	87	40	48623	5920271	90.87	91.12	91.37	
7 Shrub – tall	234	277	53	794	635	485	44	321	127	2168	1522	203	1726	85	135	621	644	10074	894832	7.36	7.88	8.41	
v8 Shrub – low	5059	2306	1736	5413	7798	3828	795	6704	1849	24419	9119	1958	5048	294	387	3799	2632	83144	8391725	9.18	9.38	9.58	
9 Herb	12289	2401	728	4394	18206	9821	281	3109	499	7854	7501	518	2913	174	108	1453	1702	73951	7873493	13.04	13.28	13.53	
13 Wetland – herb	3	0	37	3	1	5	3	140	0	29	1	0	0	0	1	1	0	224	24261	56.25	62.50	68.75	
14 Conifer – dense	2966	464	623	3313	5363	328	7862	3140	47522	313169	12047	2418	5318	50	1365	13131	1226	420305	41430973	11.21	11.31	11.40	
15 Conifer – open	1690	386	386	3575	6974	496	3917	4145	51166	240056	15843	1243	3525	72	1949	9083	1237	345743	37019733	69.28	69.43	69.59	
16 Conifer – sparse	861	390	91	1832	4952	842	185	1261	1074	10555	7491	137	889	87	272	1058	1497	33474	3684548	21.93	22.38	22.83	
17 Broadleaf – dense	795	545	41	2532	1409	265	124	511	3224	15370	9550	1553	7580	201	1149	5186	2331	52366	4884084	2.82	2.97	3.11	
18 Broadleaf – open	650	303	66	1296	1529	292	253	1050	396	4777	3245	600	3257	134	270	2124	1210	21452	2125981	14.70	15.18	15.66	
19 Broadleaf – sparse	4625	1746	141	7720	10211	2349	963	5776	916	15178	17554	1125	7513	418	339	2547	4547	83668	8889263	0.45	0.50	0.55	
20 Mixed – dense	1421	551	386	2721	2234	422	1404	2768	4298	34921	7564	1803	4977	105	827	7038	1281	74721	7669922	1.03	1.11	1.18	
21 Mixed – open	266	168	73	449	354	93	183	450	569	4944	902	290	1228	26	195	1139	216	11545	1163165	9.33	9.87	10.41	
NFI marginal	83364	14083	52012	36572	85931	22888	16885	32777	114601	721328	99401	12181	45061	1719	7142	48070	19663	1367680					
%																							
UB	0.00	31.61	85.49	2.32	9.27	43.55	0.00	0.50	41.75	33.39	7.70	13.34	7.47	26.35	12.32	2.51	0.00						
Producer	0.00	30.84	85.18	2.17	9.07	42.91	0.00	0.43	41.47	33.28	7.54	12.75	7.23	24.32	11.58	2.37	0.00						
LB	0.00	30.08	84.88	2.02	8.88	42.27	0.00	0.36	41.18	33.17	7.37	12.16	6.99	22.28	10.84	2.23	0.00						



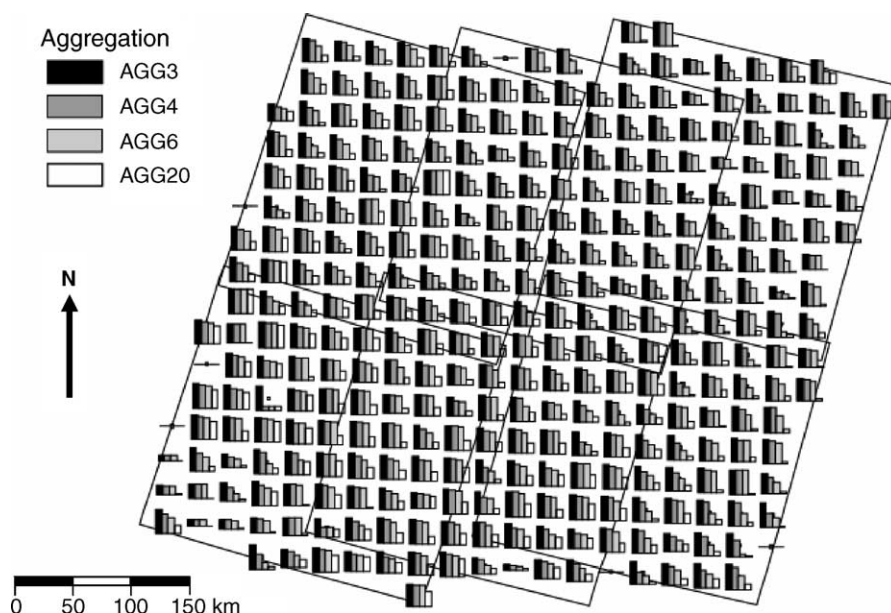


Fig. 2. Spatial distribution of overall agreement computed independently for each photo-plot and aggregation level as a series of bar graphs. The absolute heights of bars are not as interesting as their relative heights, the rate of decrease with increasing thematic detail, and the geographic variability in these patterns. Note differences among the SW corner, central-western edge, and the NE corner. In some cases all bar heights are relatively equal, in others the agreement drops regularly while some graphs exhibit very rapid agreement decreases with increasing thematic detail.

also explored. We present a series of selected maps for individual categories that illustrate their specific characteristics. These maps plot pie graphs centered on each NFI photo-plot where the gray portions of each pie represent the proportion of match between EOSD and NFI labels (by pixel), and the black portions represent mismatch. The diameter of each pie reflects the proportion of the given thematic class within the NFI photo-plot (i.e., increasingly larger pies indicate dominance by the given category).

We illustrate the different behavior of thematic classes by the following examples. When water is present in a photo-plot, the NFI matches well with EOSD, especially if it dominates the area (Fig. 3). The geographic distribution of water does not have a significant effect on the agreement of this category; there is consistent agreement between the NFI and EOSD. However, inspection of vegetated categories yields a different result. The coincidence matrices and Fig. 1 indicate the dominance of the coniferous thematic class and its prevalence in the study area. In the case of vegetation, the aggregation of categories and geographic location, significantly affect the

observed proportion of match. Fig. 4 depicts the distribution of matches for the vegetated category of AGG3, where all forested and non-forested land cover groups are combined. When forested pixels are separated from the vegetated category, slight increases in mismatch are observed, but the general agreement remains high. When the three broad forest classes are separated, geographic clusters of agreement become evident for the conifer class (Fig. 5). Clear regions are highlighted where the conifer class is dominant and where it is not, along with different proportions of match across the study area. Finally, when the dense conifer class is observed individually, the apparent geographic constraints are clear as is the decrease in agreement (Fig. 6). In this vegetation example, we see that the single, global overall agreement value could be misleading because dense conifer forests are not uniformly distributed. This demonstrates strong geographic variability and thematic aggregation effects in observed matches.

The dominance of the coniferous class within the study area dictates that a relatively high agreement could be obtained if every pixel in the study area were

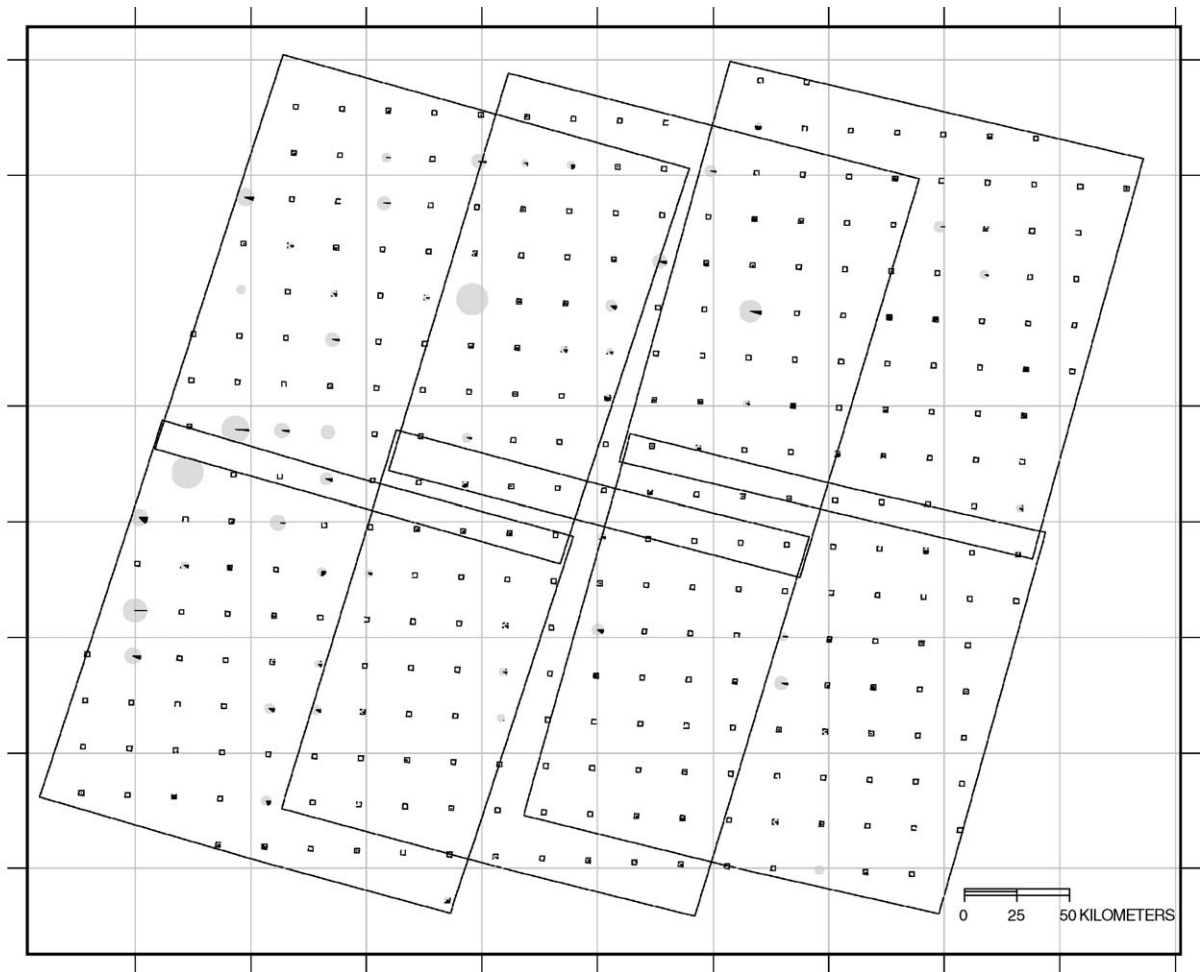


Fig. 3. Proportions of match and mismatch for the water category at AGG3 for individual NFI plots. Matches and mismatches were computed on a per-pixel basis between the NFI and EOSD data products. The gray pie wedges represent matches and the black wedges represent mismatch.

assigned “treed conifer” without even performing a valid classification. In this case, the global overall agreement values might look reasonable but the local agreement might be poor. Clearly, the Prince George study area expresses heterogeneous land cover conditions with underlying topographic and climatic gradients that influence the local vegetation. Traditional agreement assessments that are suitable for local characterization deteriorate for wide-area assessments, defining the need to stratify landscapes geographically and thematically to better typify their contributions to the actual agreement. The objective of integrating the EOSD with NFI is facilitated with realistically high coincidence due to thematic simila-

rities. However, vigilant assignment of EOSD labels is a priority to avoid making drastic mistakes regarding the incorrect allocation of thematic classes at any specific location. The exploration with pie graphs highlights how effectively we approach this objective and shows that it might be worthwhile to characterize these relationships quantitatively.

### 3.2. Entropy-based analysis

#### 3.2.1. Global agreement analysis

The off-diagonal elements of a coincidence matrix carry information about the nature of the disagreements among thematic categories (Patil and

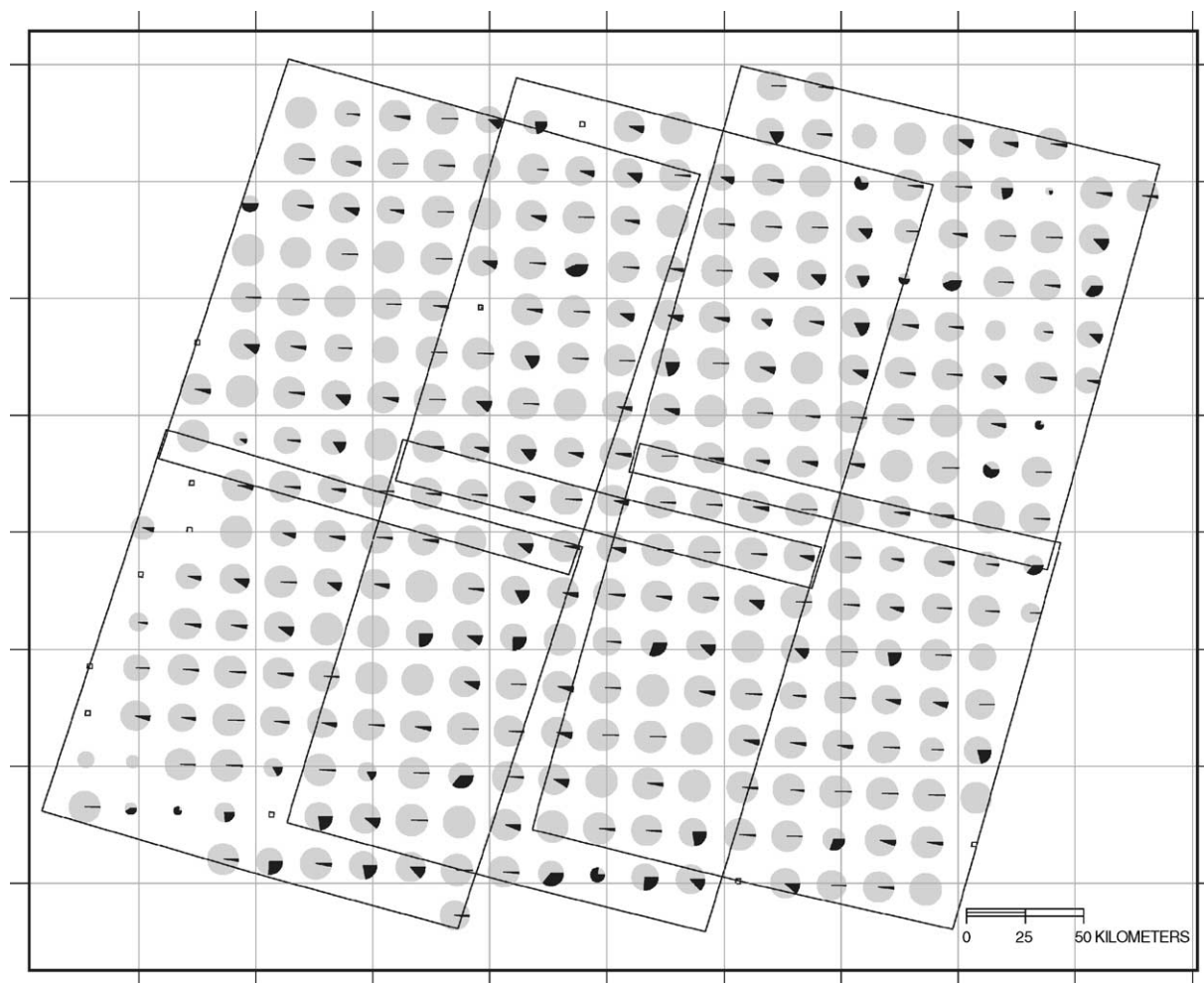


Fig. 4. Proportions of match and mismatch for the vegetated category at AGG3 for individual NFI plots. Matches and mismatches were computed on a per-pixel basis between the NFI and EOSD data products. The gray pie wedges represent matches and the black wedges represent mismatch.

Taillie, 2003). When these elements are ignored, as in the global coincidence matrix approaches, then important information might be missed for the successful integration of the NFI and EOSD products. The entropy-based analysis related to accuracy assessment (Foody, 1996) for the entire Prince George study area reveals that although water and land thematic classes are relatively insignificant in terms of occurrence, they do exhibit relatively high levels of uncertainty. The decomposition of  $U(\text{NFI}|\text{EOSD}) = 40.6\%$  among the individual thematic categories for AGG3 (computed values can be reconstructed from Table 4) illustrates significant

disagreement between land and vegetated classes, with some additional disagreement between water and land classes. The negative coefficients for land and water classes illustrate the unconstructive effect these categories have on deciphering the actual NFI category. The magnitude of these coefficients and the relative proportion of the specified thematic category indicate their overall importance on the uncertainty coefficient. Negative coefficients act to reduce the value of  $U$  and since larger values of  $U$  indicate better agreement between NFI and EOSD, it is important to interpret why negative coefficients are encountered. The decomposition for  $U(\text{NFI}|\text{EOSD}) = 40.6\% = (0.0452 \times -88.9) +$

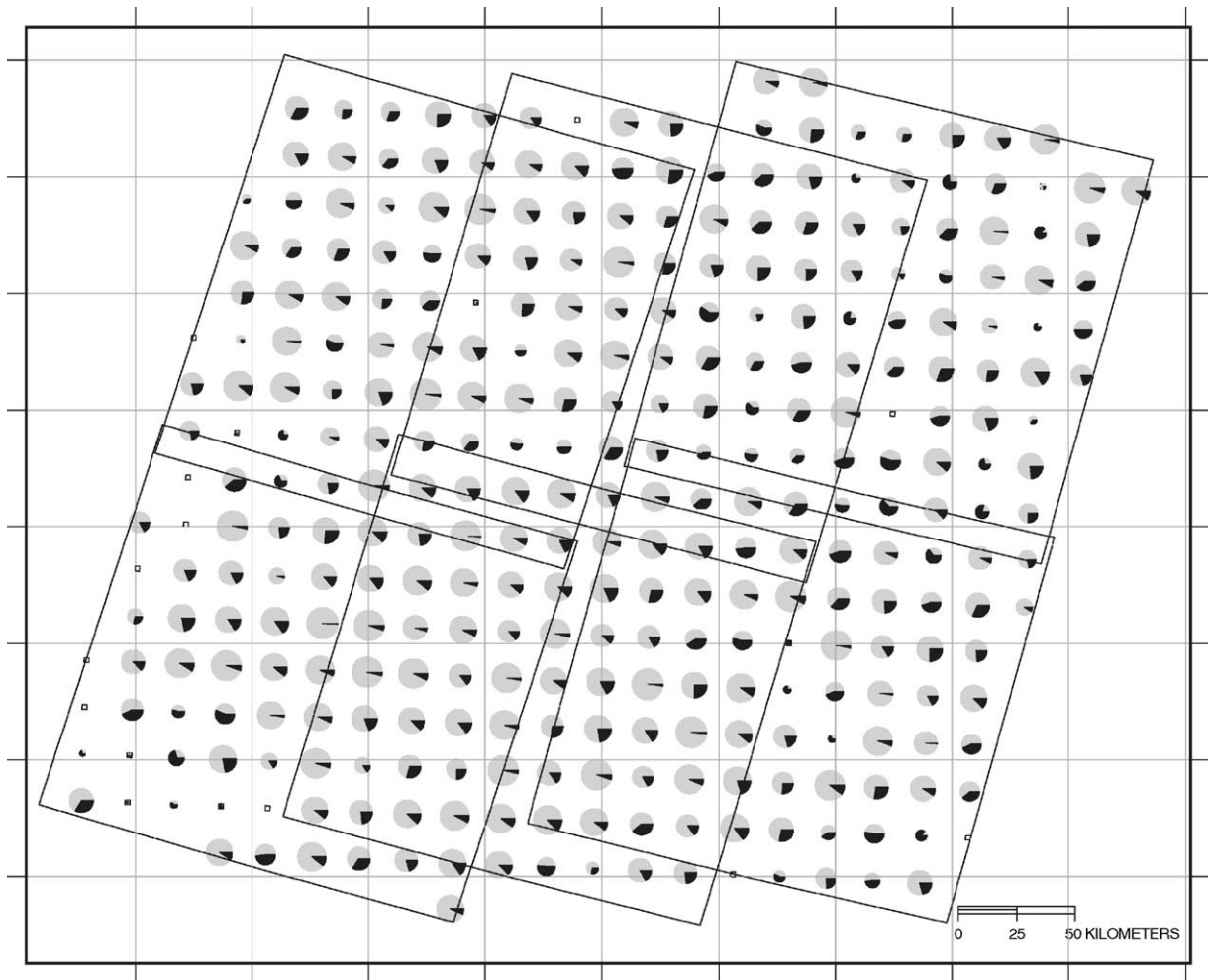


Fig. 5. Proportions of match and mismatch for the conifer category at AGG6 for individual NFI plots. Matches and mismatches were computed on a per-pixel basis between the NFI and EOSD data products. The gray pie wedges represent matches and the black wedges represent mismatch.

$(0.0696 \times -90.0) + (0.885 \times 57.4)$ , where the definition of individual terms is  $(P(\text{EOSD} = 1) \times U(\text{NFI}|\text{EOSD} = 1)) + (P(\text{EOSD} = 2) \times U(\text{NFI}|\text{EOSD} = 2)) + (P(\text{EOSD} = 3) \times U(\text{NFI}|\text{EOSD} = 3))$ , such that the proportion of each category is multiplied by its corresponding conditional uncertainty coefficient (Table 4).

The benefit of this type of analysis is the bi-directionality. We can simultaneously assess whether knowing the NFI provides significant information about the EOSD. This two-directional analysis acts directly in favour of integration because neither data set needs to be considered to be a reference. Our

illustration with AGG3 indicates that globally, for the entire Prince George study area, users will draw more benefit from knowing the EOSD before making inferences about the NFI than vice versa; this is demonstrated by  $U(\text{NFI}|\text{EOSD}) > U(\text{EOSD}|\text{NFI})$ . The decomposition for  $U(\text{EOSD}|\text{NFI}) = 37.4\% = (0.0371 \times -4.97) + (0.0661 \times -132.0) + (0.897 \times 51.6)$ .

### 3.2.2. Local agreement analysis

To illustrate the geographically decomposed effect on entropy-based measurements, we observe one of the 324 photo-plot locations at AGG3 to discuss the

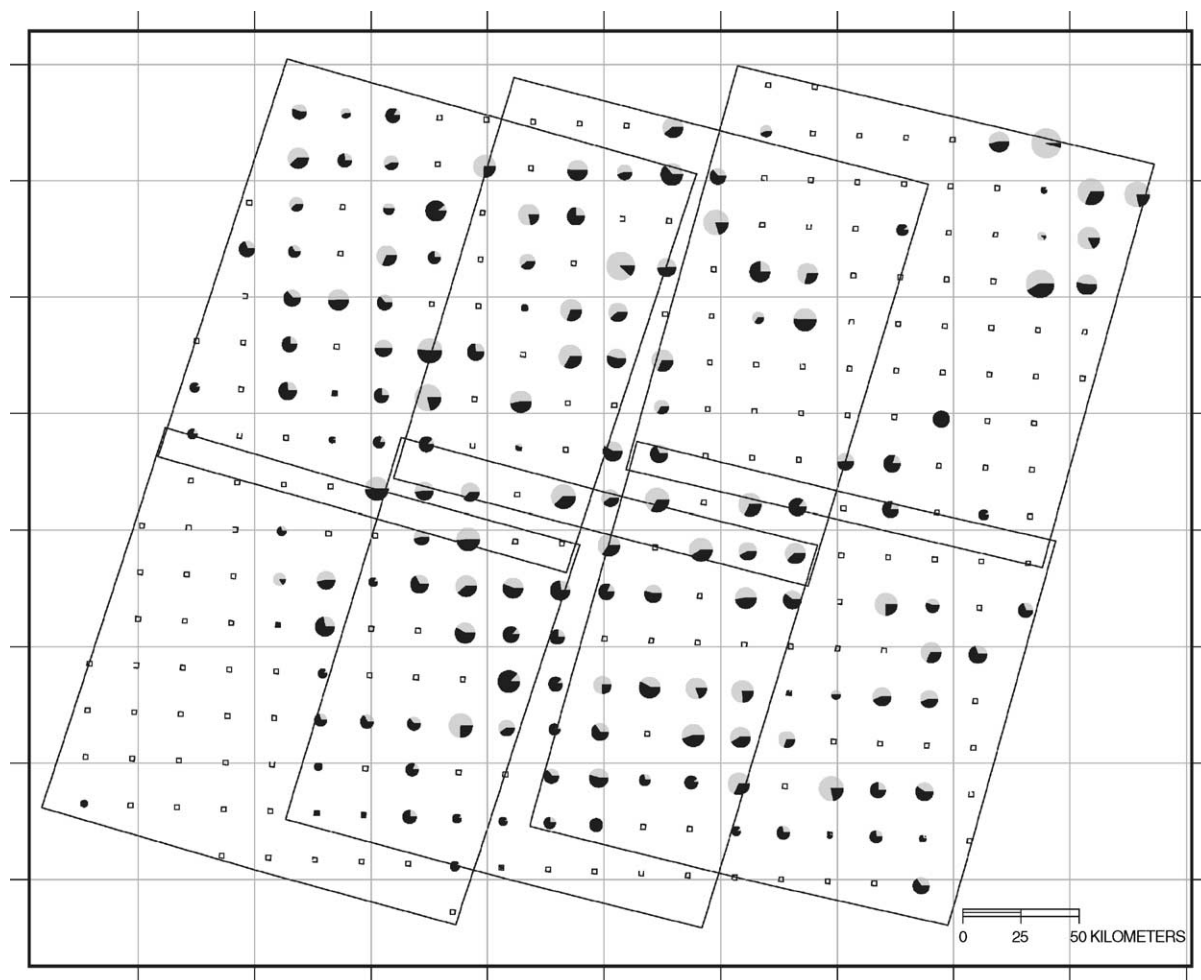


Fig. 6. Proportions of match and mismatch for the conifer-dense category at AGG20 for individual NFI plots. Matches and mismatches were computed on a per-pixel basis between the NFI and EOSD data products. The gray pie wedges represent matches and the black wedges represent mismatch.

interpretation (Fig. 7), with a local coincidence matrix (Table 8). As with the geographical decomposition of the overall agreement, local effects on the entropy-based analysis are also observed. The selected photo-plot has a slightly different relative proportion of land cover categories than the entire study area, mainly exhibiting more water and less vegetation. The overall agreement between the NFI and EOSD at this location is 91%, almost identical to the global average, but the distribution of off-diagonal elements in the coincidence matrix differs, where the only negative coefficient for uncertainty is  $U(\text{NFI}|\text{EOSD} = 2)$ . Thus, when we know the EOSD label to be land, uncertainty

is large due to disagreements between NFI categories observed in the conditional distribution. The decomposition for  $U(\text{NFI}|\text{EOSD}) = 62.6\% = (0.19 \times 42.4) + (0.0808 \times -51.5) + (0.729 \times 80.6)$ . Interestingly, the major uncertainty is found between water and land, not land and vegetation. The presence of a major meandering river through this photo-plot seems to be the culprit for causing confusion along the margins of the water body. The two-directional analysis also indicates that knowing the EOSD label greatly reduces the uncertainty of the NFI label because  $U(\text{NFI}|\text{EOSD}) > U(\text{EOSD}|\text{NFI})$ , where the decomposition for  $U(\text{EOSD}|\text{NFI}) = 57.2\% = 57.2\%$



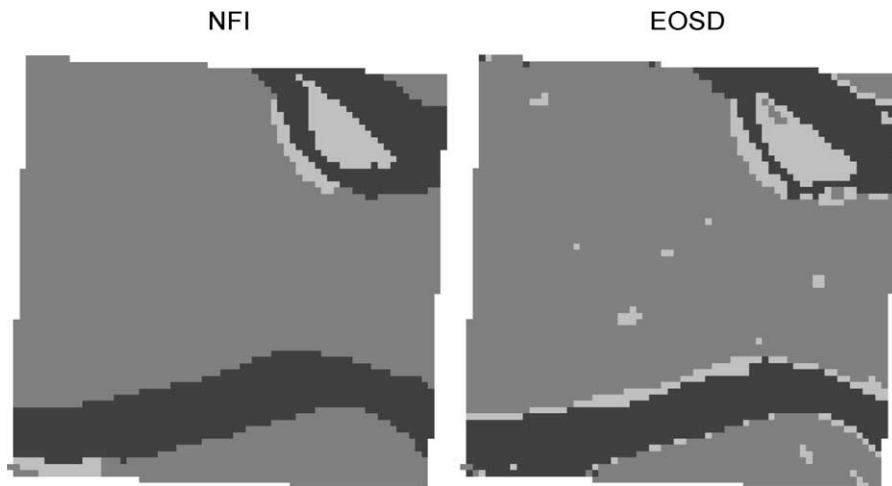


Fig. 7. Selected photo-plot exhibiting both the NFI (left) and EOSD (right) land cover classifications at AGG3. Black = water, light gray = land, dark gray = vegetation.

$(0.223 \times 7.49) + (0.0386 \times 6.63) + (0.738 \times 74.9)$ . The decomposition further allows the quantification of each thematic category's contribution to the uncertainty coefficient.

### 3.3. Data reliability

The overall agreement for a single photo-plot at each of four levels of thematic aggregation is compared to those values obtained for the core patch areas only (Fig. 8). Notice that the overall agreement at each thematic aggregation level increases as the mixed interface data are removed from the analysis. This suggests that as edges increase within an area, the expectation of agreement could decrease. It is also important to note that greater relative gains in agreement are observed for the more detailed thematic aggregations (e.g., AGG20) than for coarser ones (e.g., AGG3).

The per-pixel reliability map generated for the same photo-plot clearly indicates variability among pixels as to their distances from respective thematic cluster centroids (Fig. 9). It is obvious that the reliability of classification is not uniform across land cover types as can be seen in the variability in Fig. 9 as compared to the land cover classification in Fig. 7.

## 4. Discussion

### 4.1. Data comparison for integration

Compilation of national data products, especially for countries as large as Canada, can be difficult in terms of timing, consistency, quality, utility, and accuracy. Managing the numerous parties involved and their respective differing regional strategies can lead to fragmented or misaligned results. These data

Table 8

Local coincidence matrix for a selected photo-plot with marginal distributions as proportions and conditional uncertainty coefficient values

	NFI			EOSD proportions	$U(\text{NFI} \text{EOSD} = k)$ (%)
	Water	Land	Vegetation		
EOSD					
Water	613	22	48	0.19	42.4
Land	132	106	53	0.08	−51.5
Vegetation	58	11	2557	0.73	80.6
NFI proportions	0.22	0.04	0.74		
$U(\text{EOSD} \text{NFI} = k)$ (%)	7.49	6.63	74.9		

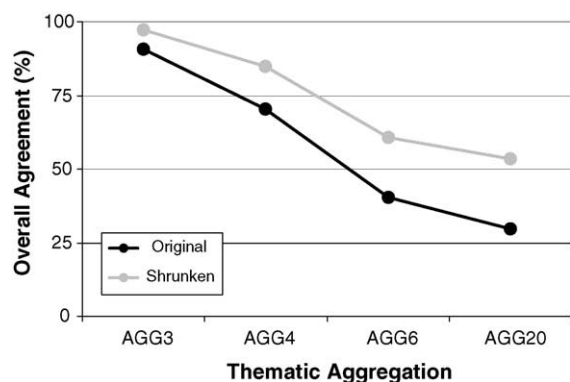


Fig. 8. Original vs. shrunk overall agreement at four levels of thematic aggregation.

and personnel management issues highlight the importance of rigorous, structured, and standardized procedures to guide the data compilation process. The NFI and EOSD were developed within the national context and thus have guidelines to maintain standards nationally, thereby simplifying their assemblage countrywide.

The efforts to standardize the EOSD classification legend greatly facilitated its integration and comparison with the existing NFI because their respective classes are related. Since EOSD is spatially continuous and the NFI represents a first-stage cluster sample, the ability to compare categories between the two data products provide the necessary response and sampling designs to conduct similarity assessments based on traditional accuracy assessment protocols to convey the agreement between the two map representations. Furthermore, the hierarchical organization of land cover labels for both the EOSD and NFI allow the comparison of these data products at multiple thematic aggregations. A distinct relationship between overall agreement and the thematic aggregation level exists, where coincidence increases dramatically at coarser thematic aggregations.

The EOSD and NFI classifications appear to agree closely for vegetated and forested land cover classes at AGG3 and AGG4; however, the agreement becomes increasingly variable and uncertain as thematic detail is increased to AGG6 and AGG20. Depending on the

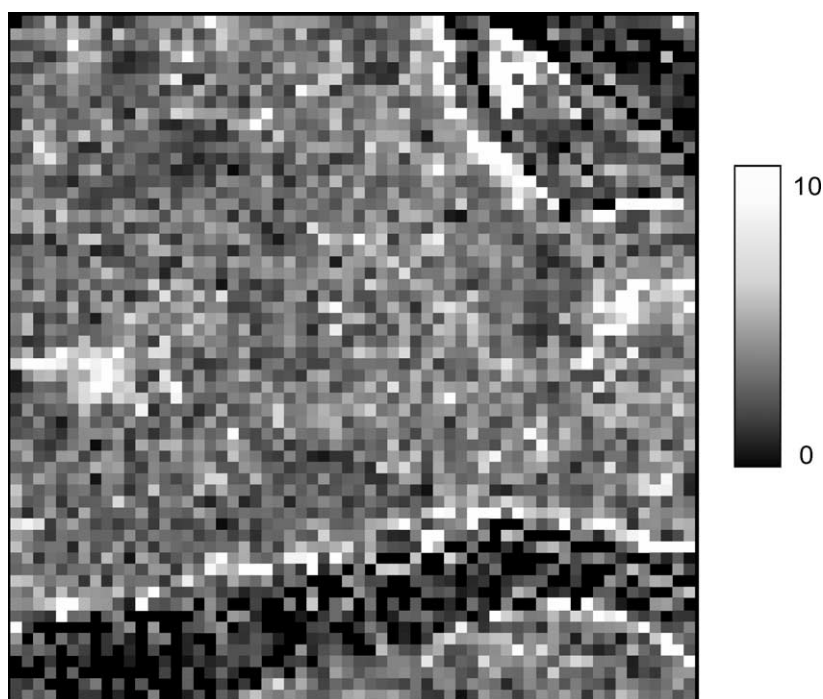


Fig. 9. Pixel reliability map for a single photo-plot. The darker pixels indicate shorter spectral Euclidean distances to the assigned category's cluster centroid, while lighter pixels indicate longer distances.

intended use, practitioners must decide what level of agreement is acceptable and then select the corresponding number of thematic classes. Since EOSD is a Canada-wide product relying on 25 m spatial resolution Landsat ETM+ imagery, local details cannot be expected to be gleaned as effectively as they might for local studies (such as when a single scene is mapped with the support of locally specific training data).

The greatest mismatches (confusion) occur among density classes (AGG20), especially for the conifer forest type. The spectral variability and definitions of dense, open, and sparse appear to be difficult to capture and differentiate among with the satellite data. The density classes are defined by crown closure percentages (>60% for dense, 26–60% for open, and 10–25% for sparse) and couple with corresponding forest types when  $\geq 75\%$  of the total basal area belongs to either the conifer, broadleaf, or mixed forest type (Wulder et al., 2003). At AGG6, confusion among the forest types (conifer, broadleaf, and mixed) is evident, but not as pronounced as the mismatch among the density classes of AGG20 because the general confusion among density classes appears confined to the given forest types. For example, mismatches in density for conifer (dense, open, and sparse) become aggregated at AGG6 and thus represent only one category called conifer. When the forest types are further aggregated to AGG4, all forest pixels, regardless of type and/or density characteristics are well matched with the NFI because the inter-forest type confusion is eliminated. Variability among forest types (and likely among ecoregions) is expected, forming the basis for potential second-stage sampling designs, where samples might be further stratified within these constraints.

Closer examination of the agreement summaries reveals a cluster of poor agreement in the southwest of the study area. The specific categories at this location reveal areas of rock and rubble that are consistently labeled by EOSD as exposed land, snow, ice, herbs, or shadow (in order of decreasing importance). The rock/rubble and exposed land categories, which are prevalent in the mountainous regions of the study area (southwest and northeast quadrants), are spectrally similar and problematic to label, resulting in a limited agreement that appears to be constrained geographically. The geographically constrained area of poor agreement identifies another possible spatial bias in need of further investigation.

#### 4.2. *New approaches toward integration*

A major development in the integration of the EOSD and NFI is to convert the coincidence matrix of counts (or proportions) to one that is entropy based (Freeman, 1987). The off-diagonal elements of a coincidence matrix carry a lot of information that is typically ignored in traditional assessments of agreement (or improperly used by the Kappa statistic when the marginals are not fixed). The conditional probability approach leads nicely into new types of questions, as opposed to the traditional “what is the coincidence between the EOSD and NFI?”, we are now able to address questions such as: “given that we know the EOSD label, what would we expect the NFI label to be, where no NFI data exists?” and “with what confidence?” Additionally, some disagreements between EOSD and NFI categories are more severe than others are (e.g., confusion among density classes for conifer versus forest types and water). These appear as off-diagonal elements and could be weighted with a cost-matrix to emphasize or suppress specific mismatches.

The spatial and thematic decompositions allow exploration into separating mismatch errors due to attributes (thematic classes) or position (Chrisman and Lester, 1991). There is also considerable interest in testing whether configuration (spatial autocorrelation) has an effect on matches/mismatches, and if there are, to define them and test their significances. Our impression is that land cover pattern influences accuracy. Intuitively, large, homogeneous land cover areas should be easier to classify correctly by the EOSD than irregular, narrow patches that potentially exhibit extensive edge effects (Zhu et al., 2000) and this is approached in our removal of interface pixels. The pixel reliability maps provide an opportunity to screen for individual pixels or clusters of unreliably classified sites such that these possible biases can be mitigated. We are also currently conducting a variety of explorations into agreement relationships with NFI polygon shapes and specific attribute data (e.g., stand age, height, crown closure) to identify further potential biases.

The traditional accuracy assessment protocols, coupled with the described extensions will not only help us to better understand the relationships between the EOSD and NFI, but will also assist with the

construction of a better sampling design. When specific categories, spatial patterns, or geographic locations are known to under perform, those specific areas, land cover types, or pattern groupings might be targeted with increased NFI sampling or ground verification efforts. The unification of all above approaches, into a usable framework could greatly enhance the information gained from an assessment of agreement and provide new opportunities for scientific inquiry. Finally, the standard support and spatial resolution can be populated with data from many sensors (especially those with finer spatial resolution), indicating the non-reliance on a single (or current) data source.

## 5. Conclusions

The primary objective of this paper was to extend traditional map comparison techniques for the two-directional assessment of agreement between the EOSD and NFI, leading towards their integration over extensive areas in Canada. This assessment was conducted globally and independently for 324 photo-plot locations at four levels of thematic detail to observe the thematic and geographic controls on the observed agreement. We demonstrated that moving beyond single number agreement summaries for large-area land cover comparisons provide substantially more information about the thematic and geographic variability within overall agreement. The results indicated that the EOSD and NFI agree reasonably well for forested areas over extensive regions, but experienced disagreement with detailed forest types and density classes. This result is not surprising as subtle spectral differences separate differing density classes. Further, non-vegetated classes such as rock/rubble and exposed land, also appear similarly in the spectral domain and exhibited some classification limitations.

The extension of the coincidence matrix approach to include off-diagonal elements through an entropy decomposition analysis provided valuable information regarding the specific contributions to the uncertainty by individual categories. While the conditional nature of these decompositions allowed for a bi-directional analysis, neither the NFI nor the EOSD ever needed to be considered as a “reference” data set.

The further ability to test the significances of these measures was also appealing from a quality control or production perspective. The geographic decomposition again illustrated differences locally from the global trend.

Finally, we explored the reliability of individual pixels by mapping their individual distances (in spectral space) from the cluster centroids representing their assigned label. These reliability maps can potentially illustrate pixels, clusters, or regions where uncertainty in the map label might prevail, encouraging further investigation or sampling to deal with the potential biases. Increased agreement following the removal of edge pixels illustrates the complex interactions among land cover interfaces, spectral signatures, and possibly landscape pattern. Further investigations are exploring these issues aimed at exposing the constraining ability of landscape element complexity and configuration. However, the currently presented approaches provide sufficient information to target unreliable locations and guide the design of further sampling to reduce the uncertainty (increase agreement) between the EOSD and the NFI.

## Acknowledgements

This research was supported in part by the GEOIDE Network of Centres of Excellence (Canada) and the Natural Science and Engineering Research Council of Canada. The authors want to thank the Canadian Forest Service for providing the data used in the analyses contained within this paper and the three anonymous reviewers that provided very useful feedback on previous versions of this manuscript.

## References

- Aronoff, S., 1982. Classification accuracy: a user approach. *Photogram. Eng. Remote Sensing* 48, 1299–1307.
- Card, D.H., 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy. *Photogram. Eng. Remote Sensing* 48, 431–439.
- Chrisman, N., Lester, M., 1991. A diagnostic-test for error in categorical maps. In: *Proceedings of Auto-Carto 10*. Vol. 6/ American Congress Surveying and Mapping, Bethesda, American Society for Photogrammetry and Remote Sensing Annual Convention, pp. 330–348.

- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing Environ.* 37, 35–46.
- Csillag, F., 1991. Resolution revisited. In: *Proceedings of Auto-Carto 10*. Volume 6/American Congress of Surveying and Mapping, Bethesda, American Society for Photogrammetry and Remote Sensing, pp. 15–28.
- Csillag, F., Boots, B., 2004. Toward comparing maps as spatial processes. In: Fisher, P. (Ed.), *Developments in Spatial Data Handling*. Springer, Heidelberg, pp. 641–652.
- Czaplewski, R.L., Patterson, P.L., 2003. Classification accuracy for stratification with remotely sensed data. *For. Sci.* 49, 402–408.
- Edwards, T.C., Moisen, G.G., Cutler, D.R., 1998. Assessing map accuracy in a remotely sensed, ecoregion-scale cover map. *Remote Sensing Environ.* 63, 73–83.
- Foody, G.M., 1996. Approaches for the production and evaluation of fuzzy land cover classifications from remotely sensed data. *Int. J. Remote Sensing* 17, 1317–1340.
- Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing Environ.* 80, 185–201.
- Freeman, D.H., 1987. *Applied Categorical Data Analysis*. Marcel-Dekker, New York.
- Gillis, M., Leckie, D., 1993. Forest inventory mapping procedures across Canada. Information Report PI-X-114. Forestry Canada, 79 pp.
- Gillis, M.D., 2001. Canada's national forest inventory (responding to current information needs). *Environ. Monit. Assess.* 67, 121–129.
- Goodman, L.A., Kruskal, W.H., 1979. *Measures of Association for Cross-classifications*. Springer-Verlag, New York.
- Gopal, S., Woodcock, C., 1994. Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogram. Eng. Remote Sensing* 60, 181–188.
- Hyypä, J., Hyypä, H., Inkinen, M., Engdahl, M., Linko, S., Zhu, Y.H., 2000. Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. *For. Ecol. Manage.* 128, 109–120.
- Kingston, J.A.E., 2004. The impact of spatial autocorrelation on confidence intervals in group (cluster) sampling for accuracy assessment. M.S.A. Thesis. University of Toronto, Toronto, Ontario, Canada, 45 pp.
- Kyriakidis, P.C., Dungan, J.L., 2001. A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions. *Environ. Ecol. Stat.* 8, 311–330.
- Mäkelä, H., Pekkarinen, A., 2001. Estimation of timber volume at the sample plot level by means of image segmentation and Landsat TM imagery. *Remote Sensing Environ.* 77, 66–75.
- Meidinger, D., Pojar, J., *Ecosystems of British Columbia*. 1991. BC Ministry of Forests Special Report Series, Victoria, BC, Canada.
- Monserud, R.A., Leemans, R., 1992. Comparing global vegetation maps with the Kappa-statistic. *Ecol. Model.* 62, 275–293.
- Natural Resources Canada, 1999. A plot-based national forest inventory design for Canada. Internal Document. Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, 70 pp.
- Natural Resources Canada, 2003. The state of Canada's forests 2002–2003. Her Majesty the Queen in Right of Canada. [http://www.nrcan-mcan.gc.ca/cfs-scf/national/what-quoi/sof/latest\\_e.html](http://www.nrcan-mcan.gc.ca/cfs-scf/national/what-quoi/sof/latest_e.html).
- Patil, G.P., Taillie, C., 2003. Modeling and interpreting the accuracy assessment error matrix for a doubly classified map. *Environ. Ecol. Stat.* 10, 357–373.
- Patil, G.P., Balbus, J., Biging, G., Jaja, J., Myers, W.L., Taillie, C., 2004. Multiscale advanced raster map analysis system: definition, design and development. *Environ. Ecol. Stat.* 11, 113–138.
- Pontius, R.G., 2000. Quantification error versus location error in comparison of categorical maps. *Photogram. Eng. Remote Sensing* 66, 1011–1016.
- Rogan, J., Miller, J., Stow, D., Franklin, J., Levien, L., Fischer, C., 2003. Land-cover change monitoring with classification trees using Landsat TM and ancillary data. *Photogram. Eng. Remote Sensing* 69, 793–804.
- Rosenfield, G.H., Fitzpatrick-Lins, K., Ling, H.S., 1982. Sampling for thematic map accuracy testing. *Photogram. Eng. Remote Sensing* 48, 131–137.
- Stehman, S.V., 1996. Use of auxiliary data to improve the precision of estimators of thematic map accuracy. *Remote Sensing Environ.* 58, 169–176.
- Stehman, S.V., 1997a. Estimating standard errors of accuracy assessment statistics under cluster sampling. *Remote Sensing Environ.* 60, 258–269.
- Stehman, S.V., 1997b. Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing Environ.* 62, 77–89.
- Stehman, S.V., 2001. Statistical rigor and practical utility in thematic map accuracy assessment. *Photogram. Eng. Remote Sensing* 67, 727–734.
- Stehman, S.V., Czaplewski, R.L., 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing Environ.* 64, 331–344.
- Stehman, S.V., Czaplewski, R.L., 2003. Introduction to special issue on map accuracy. *Environ. Ecol. Stat.* 10, 301–308.
- Wulder, M., Seemann, D., 2001. Spatially partitioning Canada with the Landsat worldwide referencing system. *Can. J. Remote Sensing* 27, 225–231.
- Wulder, M., Gillis, M., Luther, J., Dyk, A., 2001. A guide to the estimation of Canada's national forest inventory attributes from Landsat TM data. Canadian Forest Service, Pacific Forestry Centre, Working Paper, Victoria, BC, 84 pp.
- Wulder, M., Loubier, E., Richardson, D., 2002. A Landsat 7 ETM+ orthoimage coverage of Canada. *Can. J. Remote Sensing* 28, 667–671.
- Wulder, M.A., Dechka, J.A., Gillis, M.A., Luther, J.E., Hall, R.J., Beaudoin, A., Franklin, S.E., 2003. Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *For. Chron.* 79, 1075–1083.
- Wulder, M., Nelson, T., 2003. EOSD land cover classification legend report. Version 2. Canadian Forest Service/TNT Geoservices, 83 pp. [http://www.pfc.forestry.ca/eosd/cover/EOSD\\_legend\\_report-v2.pdf](http://www.pfc.forestry.ca/eosd/cover/EOSD_legend_report-v2.pdf).
- Wulder, M., Cranny, J., Dechka, J., White, J., 2004a. An illustrated methodology for land cover mapping of forests with Landsat-7



- ETM+ data: methods in support of EOSD land cover. Version 3. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, Canada, 35 pp. [http://www.pfc.forestry.ca/eosd/cover/methods\\_vers3\\_e.pdf](http://www.pfc.forestry.ca/eosd/cover/methods_vers3_e.pdf).
- Wulder, M.A., Kurz, W.A., Gillis, M., 2004b. National level forest monitoring and modeling in Canada. *Prog. Planning* 61, 365–381.
- Yang, L.M., Stehman, S.V., Smith, J.H., Wickham, J.D., 2001. Thematic accuracy of MRLC land cover for the eastern United States. *Remote Sensing Environ.* 76, 418–422.
- Zhu, Z.L., Yang, L.M., Stehman, S.V., Czaplewski, R.L., 2000. Accuracy assessment for the U.S. Geological Survey regional land-cover mapping program: New York and New Jersey region. *Photogram. Eng. Remote Sensing* 66, 1425–1435.