

Land cover harmonization using Latent Dirichlet Allocation

Zhan Li ^a, Joanne C. White ^a, Michael A. Wulder ^a, Txomin Hermosilla ^a,
Andrew M. Davidson ^{b,c} and Alexis J. Comber ^d

^aCanadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, Victoria, Canada; ^bScience and Technology Branch, Agriculture and Agri-Food Canada, Ottawa, Canada; ^cDepartment of Geography and Environmental Studies, Carleton University, Ottawa, Canada; ^dSchool of Geography, University of Leeds, Leeds, UK

ABSTRACT

Large-area land cover maps are produced to satisfy different information needs. Land cover maps having partial or complete spatial and/or temporal overlap, different legends, and varying accuracies for similar classes, are increasingly common. To address these concerns and combine two 30-m resolution land cover products, we implemented a harmonization procedure using a Latent Dirichlet Allocation (LDA) model. The LDA model used regionalized class co-occurrences from multiple maps to generate a harmonized class label for each pixel by statistically characterizing land attributes from the class co-occurrences. We evaluated multiple harmonization approaches: using the LDA model alone and in combination with more commonly used information sources for harmonization (i.e. error matrices and semantic affinity scores). The results were compared with the benchmark maps generated using simple legend crosswalks and showed that using LDA outputs with error matrices performed better and increased harmonized map overall accuracy by 6–19% for areas of disagreement between the source maps. Our results revealed the importance of error matrices to harmonization, since excluding error matrices reduced overall accuracy by 4–20%. The LDA-based harmonization approach demonstrated in this paper is quantitative, transparent, portable, and efficient at leveraging the strengths of multiple land cover maps over large areas.

ARTICLE HISTORY

Received 31 December 2019
Accepted 8 July 2020

KEYWORDS

Forest; agriculture; land cover; land use; landsat; LDA; bayesian

Introduction

Mapping land cover is critical to estimating, understanding and modeling continuous and dynamic exchanges of energy and matter between the atmosphere and the land surface (Sellers *et al.* 1997, Running 2008). Land cover mapping over large areas has evolved from manual compilations of land-related source data of varying qualities (Matthews 1983) to automated approaches that rely on Earth observation data (Townshend 1992, Wulder *et al.* 2018). Such advances have improved the efficiency, consistency, and transparency of land cover mapping over large areas, facilitating the production of land cover maps at national, continental, and global scales by various agencies (Franklin and Wulder 2002). With this new capacity comes the reality of multiple maps being generated for the same

CONTACT Joanne C. White  joanne.white@canada.ca

© 2020 Copyright of the Crown in Canada. Natural Resources Canada.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

area, driven by different information needs, agency mandates, or programmatic requirements. Therefore, the resulting maps may have discrepant land cover labels assigned to the same land areas, as reported by a range of studies comparing land cover maps (Comber *et al.* 2004a, 2004b; Giri *et al.* 2005, McCallum *et al.* 2006, Fritz and See 2008, Herold *et al.* 2008, Fritz *et al.* 2011, Kaptué Tchuenté *et al.* 2011, Pflugmacher *et al.* 2011, Pérez-Hoyos *et al.* 2012a, 2017).

Users of land cover maps may need to combine multiple overlapping maps for a given region to meet their specific objectives or information needs. Several studies have developed methods for generating new, unique land cover outputs from multiple co-existing land cover maps (Jung *et al.* 2006, Pérez-Hoyos *et al.* 2012b, Tuanmu and Jetz 2014, See *et al.* 2015, Lesiv *et al.* 2016). These studies demonstrated the benefits of multi-source land cover information through the harmonization of coexisting maps. Harmonization in the context of land cover mapping is a process that emphasizes the similarities between coexisting characterizations of land cover categories and also serves to highlight their discrepancies (Herold *et al.* 2006).

Conceptually, most harmonization methods involve casting votes to harmonized land cover (HLC) classes according to source land cover (SLC) labels. This voting process usually first establishes class associations, which are a measurement of the thematic correspondence or similarity between SLC and HLC classes. Such associations guide the votes from an SLC label to each HLC class. Specifically, the stronger the association between an SLC class and an HLC class, the more likely a label of this SLC class will vote for this HLC class. Often studies establish fixed associations using binary values (either associated or not) because they are simple and easy (Kavouras and Kokla 2002, Herold *et al.* 2008, Iwao *et al.* 2011, Tuanmu and Jetz 2014, See *et al.* 2015, Tsendbazar *et al.* 2015, 2017). Other approaches are more complex, and address the thematic overlap of class definitions by establishing soft associations using a multi- or continuous-valued variable, referred to as the semantic affinity score, between SLC and HLC classes. Semantic affinity scores are determined mostly in two ways. One approach is to follow predefined descriptive rules with subjective scoring, as exemplified by Comber *et al.* (2004a, 2004b), Jepsen and Levin (2013), Jung *et al.* (2006), and Vancutsem *et al.* (2013). Alternately, a more elaborate approach is used by calculating overlap metrics mathematically; however this approach requires detailed class definitions (Feng and Flewelling 2004, Ahlqvist 2005, Pérez-Hoyos *et al.* 2012b).

Once the associations between SLC and HLC classes are established, the SLC labels of a pixel (in the case of raster maps) vote for each HLC class accordingly. Then, votes to all the HLC classes are tallied per pixel and the HLC label of a pixel is assigned as the HLC class with the highest total number of votes cast by all its SLC labels. The total votes to an HLC class can be computed as a count of SLC labels associated with an HLC class in case of hard associations between SLC and HLC classes (Iwao *et al.* 2011), or as a sum of semantic affinity scores between all SLC labels of a pixel and an HLC class, in case of soft associations (Jung *et al.* 2006). More sophisticated approaches use a weighted count of associated SLC labels or a weighted sum of semantic affinity scores. The weights can be based on classification homogeneity (Comber *et al.* 2004b), overall classification accuracy (Jepsen and Levin 2013, Vancutsem *et al.* 2013), or complete error matrices of source maps (Tuanmu and Jetz 2014).

Given multiple maps to harmonize, we can enhance mapping confidence by demonstrating co-occurrence of compatible land cover labels from these maps. Conversely,

a lack of agreement in land cover can indicate the complexity present and assist in making decisions on harmonized land cover labels according to the different strengths and weaknesses of source maps. For example, a class in one source map may be frequently or rarely co-located with a different class in another map. Such co-occurrence of classes between maps is a manifestation of the thematic similarity and/or the confusion between classes due to classification errors. Therefore, the spatial co-occurrence of classes between source maps offers insights into semantic affinity and source map quality that may be exploited in order to harmonize maps in a quantitative, transparent, and largely automated fashion.

As large-area land cover products become more common (Wulder *et al.* 2018), quantitative and automated approaches for harmonizing land cover products will be increasingly required. In this study, we used regionalized co-occurrence information to harmonize two large-area land cover maps for southern Canada. Our objectives were two-fold. First, to develop a quantitative and data-driven harmonization approach that takes advantage of the co-occurrence information inherent in the nature of agreement and disagreement between the two maps (by incorporating both pixel-level class agreement and over-arching map-specific classification error matrices). Second, to assess and compare the harmonized maps with those generated using only semantic affinity and/or error matrices, without co-occurrence information. The aim was to develop an approach that is portable and of generic relevance beyond the data and particular maps assessed and harmonized in this study.

Study area and data

In southern Canada, a non-discrete region of transition between dominant land uses occurs (Figure 1), with complementary land cover maps resulting from mandated monitoring responsibilities respectively by Canadian Forest Service (CFS) and Agriculture and Agri-Food Canada (AAFC). The integration of the land cover maps from these agencies is desired to meet cross-sector interests and regional information needs. These two different national-level land cover products are the Virtual Land Cover Engine or VLCE from the CFS; (Hermosilla *et al.* 2018), and AAFC's Annual Crop Inventory or ACI; (Agriculture and Agri-Food Canada, 2018). These two maps have different geographic extents but overlap substantially in southern Canada at the interface of agricultural and forest dominated regions. The overlapping area is approximately 2,980,000 km² and defines the study area within which we tested our harmonization approaches for the year 2015 using VLCE and ACI as source land cover maps. Both source maps are in raster format at 30-m spatial resolution in a Lambert Conformal Conic equal-area projection, enabling a 30-m spatial resolution for harmonized map outputs.

VLCE land cover map

Annual VLCE maps were generated from gap-free best-available-pixel (BAP) composites of Landsat surface reflectance combined with auxiliary topographic data (Hermosilla *et al.* 2018). Rules were used to produce annual BAP composites free of atmospheric perturbations (clouds, cloud shadows, haze) and representing similar seasonality conditions (i.e., proximity to mid-summer target date of August 1st) (White *et al.* 2014). Then, temporal

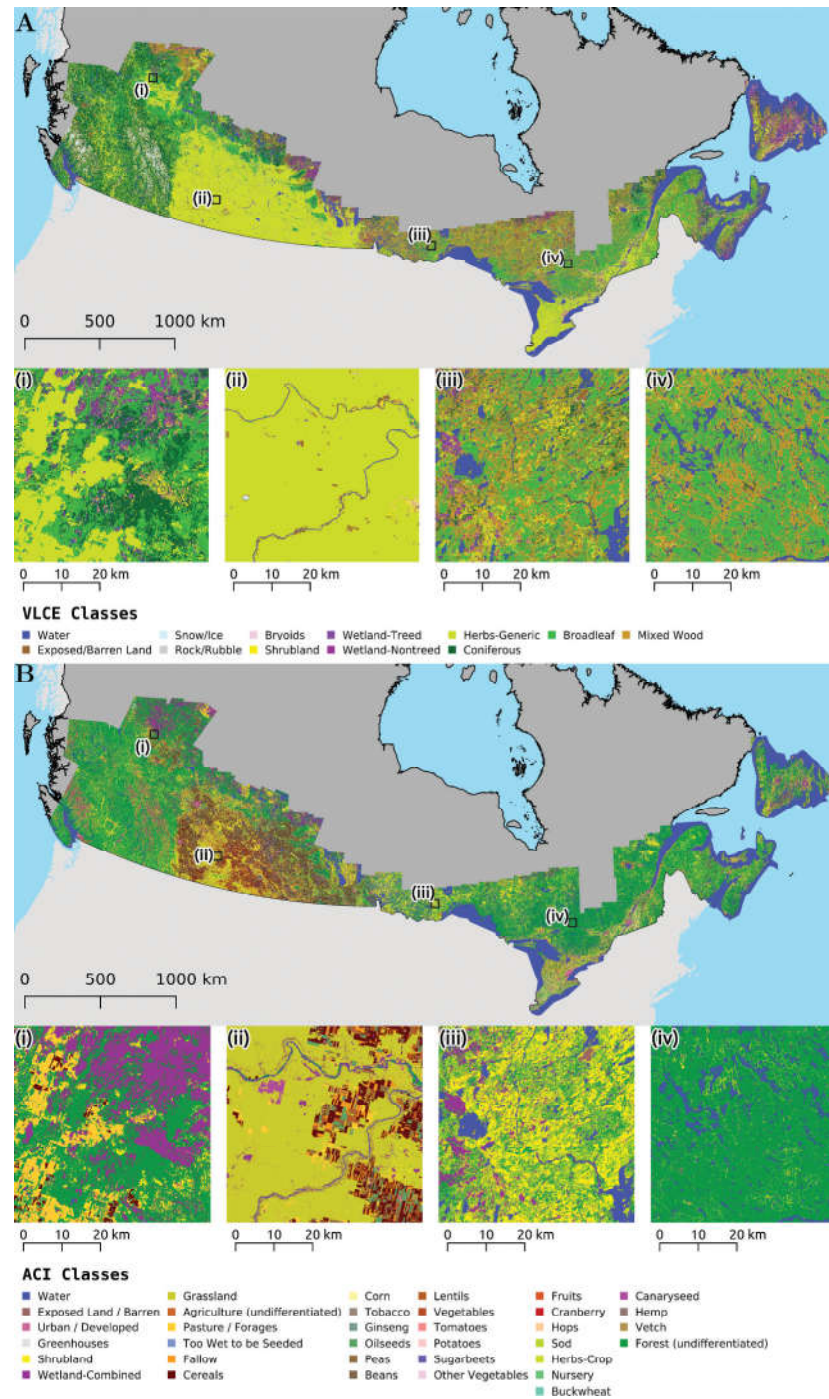


Figure 1. Study area as represented by the source maps (A) Virtual Land Cover Engine (VLCE) and (B) Annual Crop Inventory (ACI). Note that ACI map displays only the 35 ACI classes with validation samples in the accuracy assessment. ACI classes without validation samples were merged into higher level parent classes.

trend analysis was used to detect changes and to infill gaps in the BAP composites (Hermosilla *et al.* 2015). Preliminary annual land cover classifications were generated using a Random Forest classifier which were then further processed using a Hidden Markov Model (Abercrombie and Friedl 2016) with expert-based class transition probabilities to reduce the identification of spurious land cover transitions. This process generated refined annual land cover maps that are change-informed and temporally integrated (Hermosilla *et al.* 2018). The VLCE classification has twelve land cover classes including non-vegetated (i.e., water, snow/ice, rock/rubble, exposed/barren land), non-treed vegetation (i.e., bryoids, herbs, non-treed wetland, shrubland), and treed-vegetation (i.e., treed wetland, coniferous, broadleaf, mixedwood) after Wulder *et al.* (2008). The overall accuracy of the VLCE map is 70.3% ($\pm 2.5\%$ of 95% confidence interval) using an independent validation from the interpretation of high-resolution Google Earth hosted images for a representative year (2005), as described in Hermosilla *et al.* (2018).

ACI land cover map

Annual ACI maps were generated from multi-date optical and synthetic aperture radar images using a Random Forest model applied on a region-by-region basis to account for differences between adjacent scenes (Davidson *et al.* 2017). Three post-classification steps were undertaken to generate a final refined ACI product, including filtering (to reduce isolated and erroneously classified pixels), mosaicking (to properly choose class labels within overlap areas between classification regions), and adding 'permanent' classes (such as golf courses, sports fields, ski hills and airports according to a local database) (Davidson *et al.* 2017). The ACI maps use a detailed classification comprising 56 agricultural classes and 10 non-agricultural land cover classes (Agriculture and Agri-Food Canada, 2018). The overall accuracy of the 2015 ACI map is approximately 70% based on independent validation from annual crop insurance data surveys and *in situ* observations by AAFC staff (Davidson *et al.* 2017) and many of the crop classes have user's and producer's accuracies above 90%.

Methods

A generalized legend of six HLC classes (*water, non-vegetated land, shrub, wetland, herb, and treed*) was defined based on the two SLC maps and the information needs for generalized land cover classification (Wulder *et al.* 2008). The pixel by pixel co-occurrence of classes between the two maps was determined and semantic affinity scores were assigned to each pair of SLC and HLC classes (Section 3.1). Eight HLC maps were generated using different approaches under three scenarios: benchmark, simple harmonization, and complex harmonization (Table 1). In the benchmark scenario (Section 3.2), HLC maps were created directly by cross-walking the source map legends to the HLC legend. In the simple harmonization scenario (Section 3.3), two HLC maps were generated from semantic affinity scores with and without the use of error matrices. In the complex scenario (Section 3.4), the harmonization approaches analyzed co-occurrence information using Latent Dirichlet Allocation (LDA), an unsupervised statistical model developed for processing text documents, but which has been adapted to improve the semantic interoperability of land cover data (Wadsworth *et al.* 2006, Comber *et al.* 2018). Finally,

Table 1. Scenarios and approaches used to produce maps in the generalized legend of harmonized land cover classes. The coded name indicates the information used to produce the harmonized map (2 H). Information used comprise E: error matrices, L: Latent Dirichlet Allocation, S: semantic affinity scores, or alternatively ‘~’ when that information was not used. Source maps are VLCE: Virtual Land Cover Engine, and ACI: Annual Crop Inventory.

Harmonization Scenario	Coded Name	Source Maps	Information Used by Harmonization
Benchmark	VLCE2H	VLCE	Crosswalk from VLCE legend to harmonized land cover legend (Table 3)
Simple Harmonization	ACI2H	ACI	Crosswalk from ACI legend to harmonized land cover legend (Table 4)
Complex Harmonization	E~ S2H	VLCE	Error matrices
	~S2H	VLCE	~
	~L~2H	VLCE	~
	EL~2H	VLCE	Latent Dirichlet Allocation
	~LS2H	VLCE	Latent Dirichlet Allocation
	ELS2H	VLCE	Error matrices
			Semantic affinity scores
			Semantic affinity scores
			~
			~
			Semantic affinity scores
			Semantic affinity scores

we used independent validation data to assess and compare all the resulting HLC maps (Section 3.5) and to determine the reliability of harmonization using error matrices, semantic affinity scores and LDA outputs.

Comparing source land cover products and harmonizing their legends

All possible SLC classes were identified from a union of source map classes: if one VLCE class and one ACI class had the same definition, they were treated as one SLC class. Classes with the same name but different definitions were treated separately (for example, herb). The ACI classes not included in the ACI accuracy assessment were merged into their higher-level parent classes in the ACI legend. In total, 44 unique SLC classes were identified from the VLCE and ACI legends. The marginal relative frequencies of the ACI classes per VLCE class and the VLCE classes per ACI class were calculated to demonstrate the similarities and differences between the source maps. Semantic affinity scores for each of the 44 SLC classes and the 6 HLC classes were generated following an approach similar to the semantic rules of Jung *et al.* (2006). Table 2 lists the five semantic rules, along with the associated affinity scores for some example classes: higher affinity scores indicate stronger associations between two classes.

Benchmark scenario: crosswalk from VLCE/ACI legend to HLC legend

The crosswalk rules (Tables 3 and 4) were applied to link the SLC classes to the HLC classes to generate an HLC map which was validated using independent samples (Section 3.5). When both SLC classes mapped to the same HLC class, they were considered to be in agreement and in disagreement when they mapped to different HLC classes.

Simple harmonization scenario: without using LDA model

To examine the value of LDA-based harmonization, we implemented two harmonization approaches without using the LDA model: (1) by using only semantic affinity scores, and (2) by combining both error matrices of the source maps and semantic affinity scores. In the first case, a total semantic affinity score was calculated for each HLC class for a given pixel according to its SLC labels and the HLC class with the highest score was assigned to the pixel. In the second case, the summed semantic affinity scores were weighted by confusion probabilities of classes from the user's perspective based on error matrices (after Tuanmu and Jetz (2014)) to account for confusion errors in SLC labels. In both cases, if the total scores of more than one HLC class are tied, a random choice was made from these tied HLC classes.

Table 2. Definition of semantic affinity scores between example classes.

Example SLC Class	Semantic Rules	Affinity Score	Example HLC Class
Grassland	'Is not'	0	Treed
Cranberry	'Is or Has minor parts of'	1	Water
Wetland treed	'Is or Has parts of'	2	Water
Wetland non-treed			Herbs
Woody savanna (A hypothetical class)	'Is or Has major parts of'	3	Trees and grasses (A hypothetical class)
Broadleaf	'Is'	4	Treed

Table 3. Crosswalk rules from classes in the VLCE legend to classes in the HLC legend.

Original Classes in VLCE	SLC Classes	Classes in HLC Legend	Original Classes in VLCE	SLC Classes	Classes in HLC Legend
Water	Water	Water	Wetland-Treed	Wetland-Treed	Treed
Exposed/Barren Land	Exposed Land / Barren	Non-vegetated land	Wetland	Wetland-Nontreed	Wetland
Snow/Ice	Snow/Ice	Non-vegetated land	Herbs	Herbs-Generic	Herb
Rock/Rubble	Rock/Rubble	Non-vegetated land	Coniferous	Coniferous	Treed
Bryoids	Bryoids	Non-vegetated land	Broadleaf	Broadleaf	Treed
Shrubland	Shrubland	Shrub	Mixed Wood	Mixedwood	Treed

Complex harmonization scenario: using LDA model

Land cover data are discrete and comprised of categorical pixel labels with embedded semantics. Several techniques for handling discrete and categorical data and their semantics originate in the field of text mining in order to sort and classify text documents. In text mining, a text corpus is a collection of documents, themselves containing words, terms and shared topics that when considered together allow documents to be characterized as discrete and semantic data. Text mining allows for groups of documents to be statistically classified based on their component words. As such, we can assign an HLC label to a pixel based on its SLC labels in a manner that is analogous to assigning a category to a text document based on its words.

Description of Latent Dirichlet Allocation models

In text mining, Latent Dirichlet Allocation (LDA) is a well-established approach to finding representative topics in a document (Blei *et al.* 2003) that is treated as a ‘bag-of-words’ (i.e. ignoring the order of words). LDA identifies topics shared by documents in a corpus to explain the frequencies or co-occurrences of words in each document. These topics are described as probability distributions over the possible vocabulary words. In this way LDA describes each document with a set of topic probabilities, which provide a quantitative measure of a document’s content; topics with higher probabilities provide an indication of the document’s content (Blei 2012). Training an LDA model involves identifying a set of topics (i.e. a set of probability distributions over vocabulary words) that are shared across training documents to best explain the word occurrences in each document. With a trained LDA model, the most likely topic probabilities in a test document are estimated given its observed word occurrences. LDA training is a trade-off between two goals: (1) allocating as few probable topics as possible to each document; and (2) assigning as few probable words or terms as possible to each topic. Allocating few probable topics to a document requires that each topic has many terms to cover all the concepts in the document, whereas assigning few probable terms to each topic requires many more topics to be allocated to a document to cover all the concepts in the document. Training of the LDA model balances between these two goals by finding groups of tightly co-occurring words from training documents (Blei 2019).

Adapting the LDA model into land cover context

The following terminology is used in the descriptions of the LDA modeling of land cover:

Table 4. Crosswalk rules from classes in the ACI legend to classes in the HLC legend.

Original Classes in ACI	SLC Classes	Classes in HLC Legend	Original Classes in ACI	SLC Classes	Classes in HLC Legend
Water	Water	Water	Safflower	Oilseeds	Herb
Exposed Land / Barren	Exposed Land / Barren	Non-vegetated land	Sunflower	Oilseeds	Herb
Urban / Developed	Urban / Developed	Non-vegetated land	Soybeans	Oilseeds	Herb
Greenhouses	Greenhouses	Non-vegetated land	Lentils	Vegetables	Herb
Shrubland	Shrubland	Shrub	Vegetables	Beans	Herb
Wetland	Wetland-Combined	Wetland	Beans	Peas	Herb
Grassland	Grassland	Herb	Peas	Tomatoes	Herb
Agriculture (undifferentiated)	Agriculture (undifferentiated)	Herb	Tomatoes	Potatoes	Herb
Pasture / Forages	Pasture / Forages	Herb	Potatoes	Sugarbeets	Herb
Too Wet to be Seeded	Too Wet to be Seeded	Herb	Other Vegetables	Other Vegetables	Herb
Fallow	Fallow	Herb	Fruits	Fruits	Herb
Cereals	Cereals	Herb	Fruits	Fruits	Herb
Barley	Cereals	Herb	Berries	Fruits	Herb
Other Grains	Cereals	Herb	Blueberry	Fruits	Herb
Millet	Cereals	Herb	Cranberry	Cranberry	Herb
Oats	Cereals	Herb	Other Berry	Fruits	Herb
Rye	Cereals	Herb	Orchards	Fruits	Herb
Spelt	Cereals	Herb	Other Fruits	Fruits	Herb
Triticale	Cereals	Herb	Vineyards	Fruits	Herb
Wheat	Cereals	Herb	Hops	Hops	Herb
Switchgrass	Cereals	Herb	Sod	Sod	Herb
Sorghum	Cereals	Herb	Herbs	Herbs-Crop	Herb
Winter Wheat	Cereals	Herb	Nursery	Nursery	Herb
Spring Wheat	Cereals	Herb	Buckwheat	Buckwheat	Herb
Corn	Corn	Herb	Canaryseed	Canaryseed	Herb
Tobacco	Tobacco	Herb	Hemp	Hemp	Herb
Ginseng	Ginseng	Herb	Vetch	Other Crops	Herb
Oilseeds	Oilseeds	Herb	Other Crops	Other Crops	Herb
Borage	Oilseeds	Herb	Forest (undifferentiated)	Forest (undifferentiated)	Treed
Camellina	Oilseeds	Herb	Coniferous	Forest (undifferentiated)	Treed
Canola / Rapeseed	Oilseeds	Herb	Broadleaf	Forest (undifferentiated)	Treed
Flaxseed	Oilseeds	Herb	Mixedwood	Forest (undifferentiated)	Treed

- A 'document' is a collection of SLC labels in a spatially contiguous land area, for example, a tile in a raster map. An SLC label in this subset is a 'word' within this 'document'. For example, one of the pixels in a subset has SLC labels of water (VLCE) and cereals (ACI), providing two of the 'words' in this 'document'.
- A 'vocabulary' is a list of all SLC classes within the source maps.
- Latent 'topics' are the underlying concepts in a 'document' as latent land attributes in a mapped area. For example, a map tile (document) that has many SLC labels (words) of water or wetland-combined suggests it has a latent land attribute (topic) related to high wetness and the LDA would estimate high probabilities for topics containing probable words like water but low ones for topics containing probable words like urban/developed.

The LDA-based approach to harmonization

The LDA approach transforms the semantic representation of a pixel in the feature space of SLC classes (co-occurring SLC labels) to a representation in a feature space of latent land attributes (topic probabilities) that are then transformed to HLC class votes. The transformation from SLC classes to HLC is enhanced because the latent land attributes contain information of the co-occurrence of SLC classes derived from semantic similarities in class definitions and class confusions/classification errors. This LDA-based harmonization captures both shared land attributes among classes (i.e. pixel-by-pixel co-occurrence of SLC classes between the source maps) and their inherent correlations (i.e. spatial co-occurrence of SLC classes due to shared attributes among nearby pixels within the source maps). In this LDA model, a common set of latent topics (land attributes) generate co-occurring words (SLC classes) in each document (collection of labels in a contiguous mapped area) of a text corpus (entirety of mapped area). As the LDA uncovers latent topics from observed co-occurring words, we apply it to uncover latent land attributes for land cover harmonization from observed co-occurring SLC classes, without separating the two kinds of SLC class co-occurrence (i.e., pixel-by-pixel and spatial) since both are generated from shared attributes among classes.

In applying the LDA, three assumptions are made. First, within a raster map, all pixels in a map tile are similar enough to converge around a limited and small number of prominent land attributes. Second, latent land attributes derived from tiles of an appropriate size can explain the co-occurrence of SLC classes in a pixel. In other words, the spatial variations of land cover within a pixel and within a tile are statistically comparable over many pixels and tiles. Third, topics (land attributes) of a pixel that determine its HLC class label are latent and have no explicit meaning. Topic probabilities are converted into HLC class votes by assigning each topic to an HLC class, following the assumption that each topic (land attribute) is an indication of one HLC class and each HLC class may manifest as multiple topics (land attributes).

Considering these three assumptions, we applied the following steps to determine HLC labels in the LDA-based harmonization:

- (1) Training an LDA model with documents (spatially-contiguous areas) constructed by simple tiles of a chosen size to discover latent topics (land attributes) (Section 3.4.4).
- (2) Estimating probabilities of latent topics (land attributes) from the co-occurrence of SLC classes in each pixel via the trained LDA model (Section 3.4.5).

- (3) Converting topic probabilities to HLC class votes by assigning latent topics to the HLC classes in our target legend (Section 3.4.6).

Training an LDA model with tiles of source maps

We selected the tile size for training that yielded the best LDA performance to predict SLC labels in individual pixels. We trained the LDA over all the tiles in the study area and evaluated the LDA performance by a standard measure called perplexity (Blei *et al.* 2003, Griffiths and Steyvers 2004) over randomly selected test pixels (20% of pixels in the study area). Perplexity indicates the uncertainty in predicting a single word in test documents by a trained LDA model with its discovered topics; lower values are better, and chance performance results in a perplexity equal to the size of the vocabulary (Griffiths and Steyvers 2004). Using the above criteria, we determined a tile size of 300×300 pixels ($\sim 81 \text{ km}^2$). We also needed to prescribe the number of topics for the LDA training. The perplexity was minimized (and stable) as the topic count increased to 72 and therefore, we used 72 topics in the LDA model training.

We used an implementation of the LDA model in the Scikit-learn Python package (Pedregosa *et al.* 2011). The LDA model training takes the input of a document-word matrix that is built from a collection of documents by populating each row with the frequencies of vocabulary words per document. In this analysis, a row in this matrix becomes the frequencies of SLC classes given the labels from each tile of 300×300 pixels. We constructed the document-word matrix from tiles of the source maps in two ways, the first without using error matrices of source maps and the second using error matrices. Essentially, without using error matrices, the frequencies of SLC classes in a map tile (i.e., a row of the document-word matrix) are simple sums of the frequencies from both source maps. When using error matrices, we populated the document-word matrix by classification-error-adjusted frequencies of SLC classes in map tiles. Adjusting frequencies per map tile essentially resamples SLC labels and redistributes their frequencies according to their classification errors respectively for the source maps in a similar fashion as in Tuanmu and Jetz (2014).

Estimating probabilities of latent topics underlying source land cover labels

After training the LDA model, we estimated the topic probabilities per pixel that are latent but quantitative representations of land attributes given by its SLC labels. The trained LDA can take in a sequence of enough SLC labels, i.e. a document, to estimate topic probabilities underlying these labels. While two SLC labels per pixel from the source maps are too scarce for a trained LDA model to estimate reliable topic probabilities, we generated a pseudo document of the average size of label (word; W) counts (\bar{N}_W) per map tile to estimate topic probabilities given a combination of one VLCE and one ACI label. The purpose of the pseudo document is to enable the trained LDA model to estimate topic probabilities when given only two SLC labels (one from VLCE and one from ACI). Preliminary investigations indicated that the LDA is sensitive to document sizes, in this case class descriptions, as also found by Wadsworth *et al.* (2006). To avoid inconsistencies due to document sizes, a trained LDA model should be applied to documents of similar sizes to those documents on which it was trained. Therefore, the purpose of using pseudo documents of the same size as the training tiles is to obtain reliable estimates of topic

probabilities while avoiding inconsistencies due to large differences in document sizes between training and topic estimation.

Without using an error matrix, per combination of a VLCE label i_v and an ACI label i_a , each label is simply replicated for $\frac{\bar{N}_{Wj}}{2}$ times to formulate a pseudo document. Such a replication preserves the probabilities of latent topics (land attributes) underlying the original two SLC labels per pixel. When using error matrices, we adjusted the replication times per class based on classification errors to penalize the probabilities of topics related to less accurate SLC classes. Specifically, we replicated the label i_v into a label of class j for $\frac{\bar{N}_{Wj}}{2} \cdot \hat{u}_{i_vj}$ times, where \hat{u}_{i_vj} is the estimated probability that a pixel mapped as the class i_v in the VLCE map has true land cover of the class j . Similarly, we replicated the label i_a into a label of class j for $\frac{\bar{N}_{Wj}}{2} \cdot \hat{u}_{i_a j}$ times, where $\hat{u}_{i_a j}$ means the same as \hat{u}_{i_vj} but for the ACI map. We estimated \hat{u}_{i_vj} and $\hat{u}_{i_a j}$ from the error matrices of the source maps similarly as in Tuanmu and Jetz (2014).

Converting latent topic probabilities into votes to harmonized land cover classes

We converted topic probabilities into HLC class votes by assigning each topic to an HLC class. After estimating the probabilities of topics given a combination of two SLC labels respectively from source maps, we calculated the total probability of topics that we assigned to an HLC class. We used this total probability as the vote to this HLC class for a pixel and then labeled this pixel with the HLC class of the highest vote.

We assigned the topics to HLC classes by using semantic affinity scores, and without using these scores. Essentially, when using semantic affinity scores, we first identified the two most important SLC classes to a topic for the assignment. A trained LDA model quantified the importance of SLC classes to a topic by estimating the conditional probabilities of SLC classes for a topic. Higher probabilities indicate more importance. We chose the top two classes per topic since each pixel contains information from two labels. Per topic, we calculated a total semantic affinity score for each HLC class from its two most important SLC classes weighted by their conditional probabilities for this topic. Then we assigned this topic to the HLC class with the highest weighted sum of scores.

When assigning topics without using semantic affinity scores, we took advantage of the combinations of agreed classes between the source maps. For each of the 46 combinations in agreement (grid cells with black dots in Figures 2 and 3), its two SLC classes corresponded to the same HLC class according to the direct crosswalk from SLC legends to the HLC legend (Tables 3 and 4). Each combination in agreement therefore corresponded to this HLC class shared by its two SLC classes. Given any pixel from any combination in agreement, its votes to HLC classes are assumed to be 1 for its corresponding HLC class and 0 for the others. Using the known HLC votes of pixels from combinations in agreement and their topic probabilities as training data, we found a set of correspondences between topics and HLC classes, i.e. assignment of topics to HLC classes, by minimizing squared sum of differences between the predicted and known votes to HLC classes for the combinations in agreement.

Quantifying confidence in harmonized land cover labels

As the exact estimation of topic probabilities for LDA is intractable (Blei *et al.* 2003), only approximate estimation is available. Therefore, multiple runs of LDA model training and

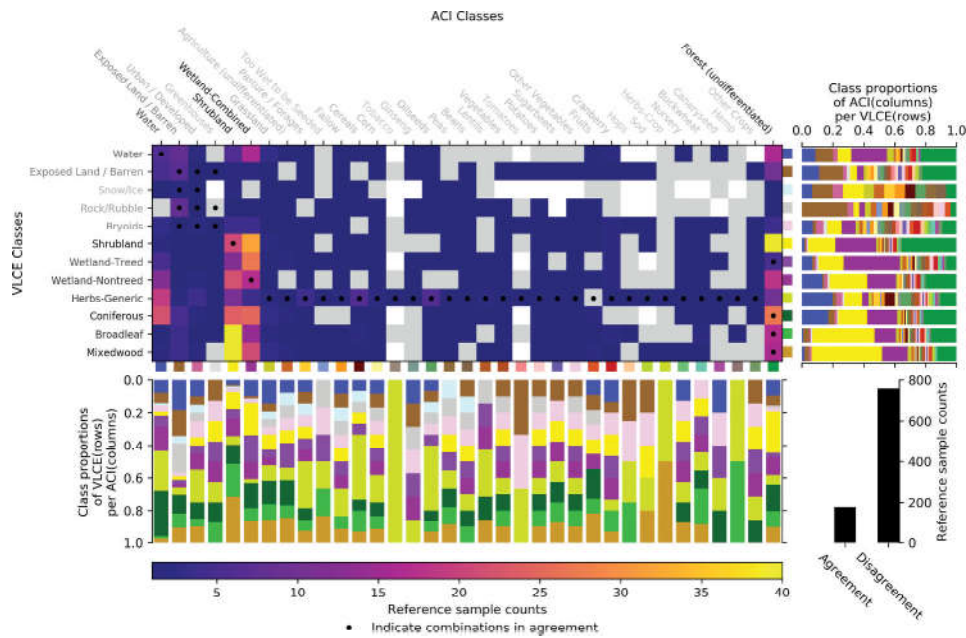


Figure 2. Reference sample allocation. In the matrix image on the upper left, blanks indicate no such combination and grays indicate no reference sample is allocated. Darker fonts of class names mean higher frequencies. The bar charts on the right of and directly below the image are marginal proportions. Class colors are presented on the right Y-axis and the bottom X-axis of the matrix image. The proportion of reference sample units over areas of agreement and disagreement is presented on the lower-right bar chart.

estimation may yield different results of harmonization because these approximate algorithms may find different latent topics (different probability distributions over vocabulary words) and hence estimate different topic probabilities each time. The randomness due to different results of discovered latent topics offers a chance to quantify the confidence in estimated HLC labels per combination of source map classes. We ran LDA model training for 400 iterations and bootstrapped these model outputs to 4000 resamples, from which we found the most frequent HLC labels and calculated their relative frequencies as class probabilities to quantify the confidence in HLC labels. Intuitively, the more likely a combination is to occur, the more consistent the LDA outputs will be and hence the higher the class probabilities. For example, the co-occurrence of VLCE water and ACI water is common, leading to higher confidence in the estimated HLC label (water) than other less likely combinations, such as water by VLCE and shrubland by ACI.

Assessing and comparing approaches to land cover harmonization

All harmonized land cover outputs were assessed using independent validation sample pixels. An overall sample size of 1200 was selected assuming an overall map accuracy of 0.85 and a targeted margin of error of 0.02 for 95% confidence interval using the following equation (Cochran 1977; Olofsson *et al.* 2014),

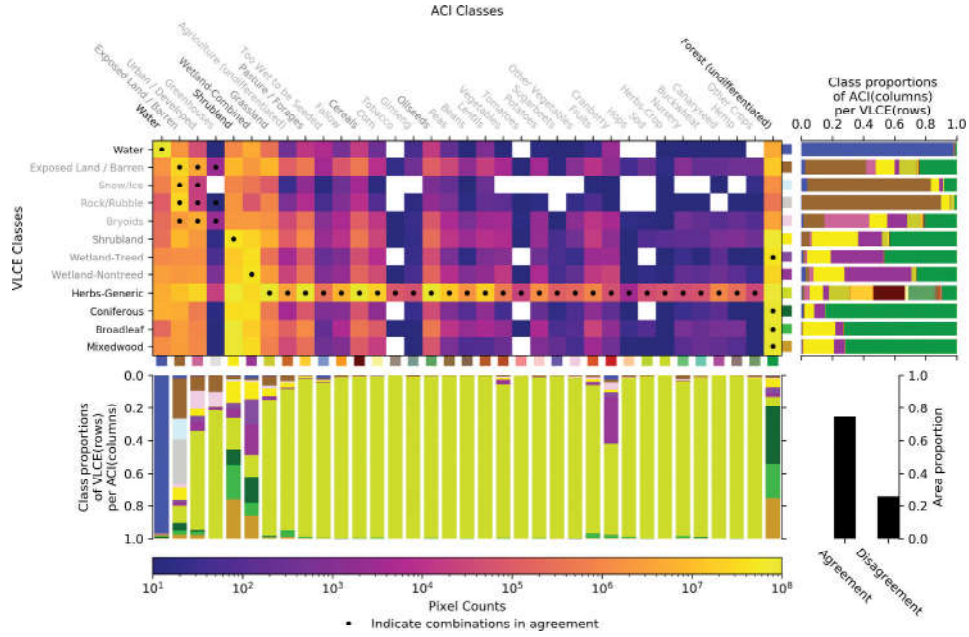


Figure 3. Co-occurrence frequencies (pixel counts) of combinations of source map classes. In the matrix image on the upper left, blanks indicate no such combination. Darker fonts of class names mean higher frequencies. The bar charts on the right of and directly below the image are marginal proportions. Class colors are presented on the right Y-axis and the bottom X-axis of the matrix image. The proportion of combinations in agreement and disagreement is presented on the lower-right bar chart.

$$n = \left(\frac{z}{m}\right)^2 \cdot a \cdot (1 - a)$$

where a is conjectured accuracy expressed as an areal proportion, z is the percentile from the standard normal distribution (1.96 for 95% confidence interval), and m is the margin of error.

We designed a two-step hierarchical stratified random sampling approach to allocate validation samples to each of the HLC class strata and then to each combination of source map classes (Figure 2). This sampling strategy balanced the need to sample those combinations of VLCE and ACI classes that were more common, with those combinations of VLCE and ACI classes that were likely to result in less accurate HLC labels in areas of disagreement. Sample pixels were allocated to HLC classes according to their expected accuracies and their average mapped areas, as derived from the HLC maps VLCE2H and ACI2H. As indicated in Figure 2, more samples were allocated to areas of disagreement.

High-resolution images on Google Earth were interpreted to assign a primary label to each sample pixel (with a 30 m \times 30 m area as a spatial support region, same as the resolution of source maps). We also assigned a secondary label to each sample pixel to accommodate both thematic ambiguity and spatial accuracy in land cover mapping (Stehman and Czaplewski 2003). After the visual interpretation, we excluded sample pixels that had no available high-resolution images on Google Earth and obtained in total 936 validation sample pixels as reference for accuracy assessment. Agreement

between map and reference data was defined if a map label matched either primary or secondary reference label. We summarized accuracy assessment results for each harmonization approach in error matrices that were estimated using the reference data and mapped class proportions following Olofsson *et al.* (2014). We then computed the overall accuracy, user's and producer's accuracy per class, and their standard error from the error matrix of a map over areas of agreement, and disagreement respectively.

Over areas of disagreement, land cover labels from one source map may be correct, or labels from both source maps may be incorrect. The best a harmonization approach can achieve is to either choose the winning HLC labels given by source maps, or wisely choose an HLC class different from that given by any source map when it detects both source labels are wrong. If a harmonization approach can achieve the above ideal performance for all the pixels, it will result in the best possible harmonized map that has higher values for all the accuracy quantities (overall, user's and producer's) than any source map. However, it is difficult, if not impossible, to automatically achieve this ideal harmonization. Therefore, over areas of disagreement, a good result from a harmonization approach gives higher accuracy than all source maps, whereas a fair result gives accuracies that are somewhere in between source maps. Conversely, a poor harmonization would result in an accuracy that is lower than all the source maps. The use of these criteria to identify the best overall performance of harmonized maps was corroborated by previous studies on land cover harmonization, most of which reported overall, and user's and producer's accuracies of harmonized maps exceeding at least one source map, rarely being lower than all source maps, but also rarely exceeding all source maps (Iwao *et al.* 2011, Pérez-Hoyos *et al.* 2012b, Kinoshita *et al.* 2014, Tuanmu and Jetz 2014, Tsendbazar *et al.* 2017).

Results

Comparison of source land cover maps

We found the source map (VLCE and ACI) land cover classes agreed with each other for more than 75% of the total map area (Figure 3, combinations in agreement are marked by black dots in the co-occurrence matrix). Thus, despite being produced by two separate agencies with different mapping objectives, the source land cover maps achieved a substantial area of agreement for land cover classification. Our harmonization efforts therefore focused on reconciling disagreement in the remaining 25% of the total map area. The total area of agreement indicates that the co-occurrence of agreed classes between source maps is not by chance and provides increased confidence in the classification outcomes for pixels labeled with highly agreed classes. However, across different source map classes, the marginal proportions of pixels with agreed labels are uneven. For example, more than 90% of water pixels in VLCE agreed with ACI and vice versa while less than 40% of shrubland pixels agreed. Although most agricultural classes in the ACI map agree with the VLCE herb class over 90% of their classified area, pixels classified as ACI cranberry have much lower agreement with the VLCE herb class, and stronger agreement with the VLCE wetland non-treed class. Different levels of agreement and disagreement imply varying semantic similarities between source map classes as well as different levels of class confusions, both of which are a manifestation of shared land attributes among classes and are informative to the harmonization.

Comparison of complex and simple harmonization approaches

We found almost identical estimates of accuracy over areas of agreement (Table 5) for all the maps that were generated with the different harmonization approaches (Table 1), all of which fall within the accuracy constraints set by the two benchmark land cover maps ('VLCE2H' and 'ACI2H'). Over these areas, the overall accuracy estimates for the harmonized maps were close to 0.9. Both the user's and producer's accuracy estimates for most HLC classes also achieved high values around or above 0.8, except for shrubland and wetland classes. In contrast, over areas of disagreement, there was variability in accuracy measures of harmonized maps according to the approach applied (Table 6).

With the evaluation criteria stated in Section 3.5 in mind, we found that the following three harmonization approaches had better overall performance for areas of disagreement between the VLCE and ACI maps: 'E~2SH', 'EL~2H', and 'ELS2H'. All three of these approaches use error matrices. The harmonized maps from the two approaches using the LDA (Figure 4) are quite similar, with some differences that we explain in more detail in Section 4.3. The increase in overall accuracy relative to the VLCE-based benchmark map (VLCE2H) ranged from 4–10% and about 7–23% compared to the ACI benchmark map (ACI2H). 'ELS2H' was the only harmonized map that yielded both user's and producer's accuracies for all classes that exceeded the accuracy of at least one benchmark map (Table 6).

Efficacy of error matrix, LDA model and semantic affinity for the harmonization

In our assessment, we separately tested the utility of three information sources for harmonization: error matrices, LDA outputs, and semantic affinity scores. We used the change in accuracy of the resulting output maps when excluding an information source as an indicator of its value for harmonization. The exclusion of error matrices resulted in the greatest decrease in accuracy measures (Table 7), with the change of overall accuracy ranging from -4% to -20%, a median change in producer's accuracy of -5%, and a median change in user's accuracy of -3%, highlighting the importance of error matrices in land cover harmonization. In contrast, excluding the rank-based

Table 5. Estimates of overall, producer's (P_i), and user's (U_i) accuracy per HLC class over areas of agreement, with standard error. Codes for harmonization scenarios are fully described in Table 1.

Area of Agreement	VLCE2H	ACI2H	~S2H	E~S2H	~L~2H	~LS2H	EL~2H	ELS2H
Producer's	Overall	.89 ± .04	.89 ± .04	.89 ± .04	.89 ± .04	.89 ± .04	.89 ± .04	.88 ± .04
	Water	.91 ± .05	.91 ± .05	.91 ± .05	.91 ± .05	.91 ± .05	.91 ± .05	.92 ± .05
	Non-veg.	.79 ± .13	.79 ± .13	.79 ± .13	.79 ± .13	.78 ± .13	.79 ± .13	.78 ± .13
	Shrub	.37 ± .13	.37 ± .13	.37 ± .13	.37 ± .13	.37 ± .13	.37 ± .13	.37 ± .13
	Wetland	.79 ± .13	.79 ± .13	.79 ± .13	.79 ± .13	.79 ± .12	.79 ± .13	.79 ± .13
	Herbs	.97 ± .03	.97 ± .03	.97 ± .03	.97 ± .03	.94 ± .04	.97 ± .03	.92 ± .04
User's	Treed	.89 ± .07	.89 ± .07	.89 ± .07	.89 ± .07	.89 ± .07	.89 ± .07	.89 ± .07
	Water	.83 ± .17	.83 ± .17	.83 ± .17	.83 ± .17	.83 ± .17	.83 ± .17	.83 ± .17
	Non-veg.	.84 ± .07	.84 ± .07	.84 ± .07	.84 ± .07	.85 ± .07	.84 ± .07	.84 ± .07
	Shrub	.62 ± .11	.62 ± .11	.62 ± .11	.62 ± .11	.62 ± .11	.62 ± .11	.62 ± .11
	Wetland	.47 ± .12	.47 ± .12	.47 ± .12	.47 ± .12	.47 ± .12	.47 ± .12	.47 ± .12
	Herbs	.95 ± .03	.95 ± .03	.95 ± .03	.95 ± .03	.95 ± .03	.95 ± .03	.95 ± .03
	Treed	.92 ± .03	.92 ± .03	.92 ± .03	.92 ± .03	.91 ± .04	.92 ± .03	.90 ± .04

Table 6. Estimates of overall, producer's (P_j), and user's (U_i) accuracy per HLC class over areas of disagreement, with standard error. Codes for harmonization scenarios are fully described in Table 1.

Area of Disagreement		VLCE2H	ACI2H	~S2H	E~S2H	~L~2H	~LS2H	EL~2H	ELS2H
Producer's	Overall	.52 ± .02	.39 ± .02	.42 ± .02	.62 ± .02	.54 ± .02	.51 ± .02	.58 ± .02	.56 ± .02
	Water	.16 ± .03	.13 ± .02	.10 ± .01	.15 ± .02	.00 ± .00	.00 ± .00	.26 ± .03	.18 ± .02
	Non-veg.	.39 ± .06	.33 ± .07	.67 ± .05	.51 ± .06	.47 ± .06	.52 ± .06	.54 ± .06	.47 ± .06
	Shrub	.46 ± .07	.68 ± .05	.86 ± .03	.34 ± .06	.14 ± .05	.12 ± .05	.44 ± .06	.47 ± .06
	Wetland	.15 ± .03	.62 ± .05	.22 ± .03	.28 ± .04	.33 ± .04	.60 ± .05	.22 ± .04	.40 ± .04
	Herbs	.57 ± .05	.25 ± .03	.67 ± .04	.78 ± .04	.74 ± .03	.71 ± .04	.77 ± .04	.69 ± .04
User's	Treed	.70 ± .03	.32 ± .02	.23 ± .02	.76 ± .02	.70 ± .02	.59 ± .02	.67 ± .02	.65 ± .02
	Water	.79 ± .05	.53 ± .06	.75 ± .07	.70 ± .06	1.0 ± .00	.00 ± .00	.71 ± .05	.79 ± .05
	Non-veg.	.48 ± .05	.28 ± .07	.46 ± .04	.33 ± .05	.33 ± .06	.35 ± .05	.43 ± .07	.50 ± .08
	Shrub	.30 ± .05	.23 ± .03	.27 ± .03	.40 ± .06	.17 ± .06	.16 ± .07	.33 ± .05	.34 ± .05
	Wetland	.23 ± .05	.36 ± .04	.59 ± .10	.61 ± .09	.33 ± .05	.37 ± .04	.33 ± .06	.41 ± .05
	Herbs	.45 ± .07	.56 ± .03	.55 ± .04	.54 ± .03	.53 ± .03	.52 ± .03	.55 ± .03	.54 ± .03
	Treed	.68 ± .02	.65 ± .05	.63 ± .04	.75 ± .03	.68 ± .03	.67 ± .03	.73 ± .03	.68 ± .03

semantic affinity scores from the harmonization resulted in a small increase in overall accuracy of 2% to 3%, with a median change in producer's accuracy of 2% and no change in user's accuracy (Table 7). Compared to error matrices and semantic affinity scores, the effect of excluding the LDA model was mixed. When LDA models were excluded from the ~LS2H scenario (which does not include an error matrix; Table 1), overall accuracy decreased by 10%, with a median change in producer's accuracy of 3% and a median change in user's accuracy of 11%. For the ELS2H scenario however, excluding the LDA from the harmonization resulted in an increase in overall accuracy of 5%, with no change in producer's accuracy and a median change in user's accuracy of 3%.

Confidence in harmonized land cover labels and quality of harmonization

We visualized as matrices the harmonization results for the 'EL~2H' (Figure 5) and 'ELS2H' (Figure 6) approaches. The spatial details of the harmonization results (both HLC classes and class probabilities) in comparison with the benchmark maps are demonstrated over an example region near the west of Calgary, Alberta (Figure 7). As expected, the class combinations in agreement generally have high confidence in HLC labels (Figures 5 and 6). In the 'ELS2H' (Figure 6), the following two combinations in agreement do not follow their SLC labels from VLCE and ACI, (1) VLCE herbs and ACI cranberry, and (2) VLCE herbs and ACI buckwheat. However, the class probabilities for both combinations are low, suggesting caution when using the harmonized labels of these two combinations. For those intuitively less probable combinations that have thematically distant classes (e.g., VLCE shrubland and ACI wetland-combined), their low class probabilities also indicate cautions with their HLC labels. Such cases may result from serious commission errors of SLC classes from the source maps, which suggests a need for a manual review of those areas in the source maps to determine the most appropriate HLC label. Therefore, the LDA-based harmonization approaches allow the generation of useful class probabilities via multiple runs of LDA models that are essential for interpreting the harmonized maps.

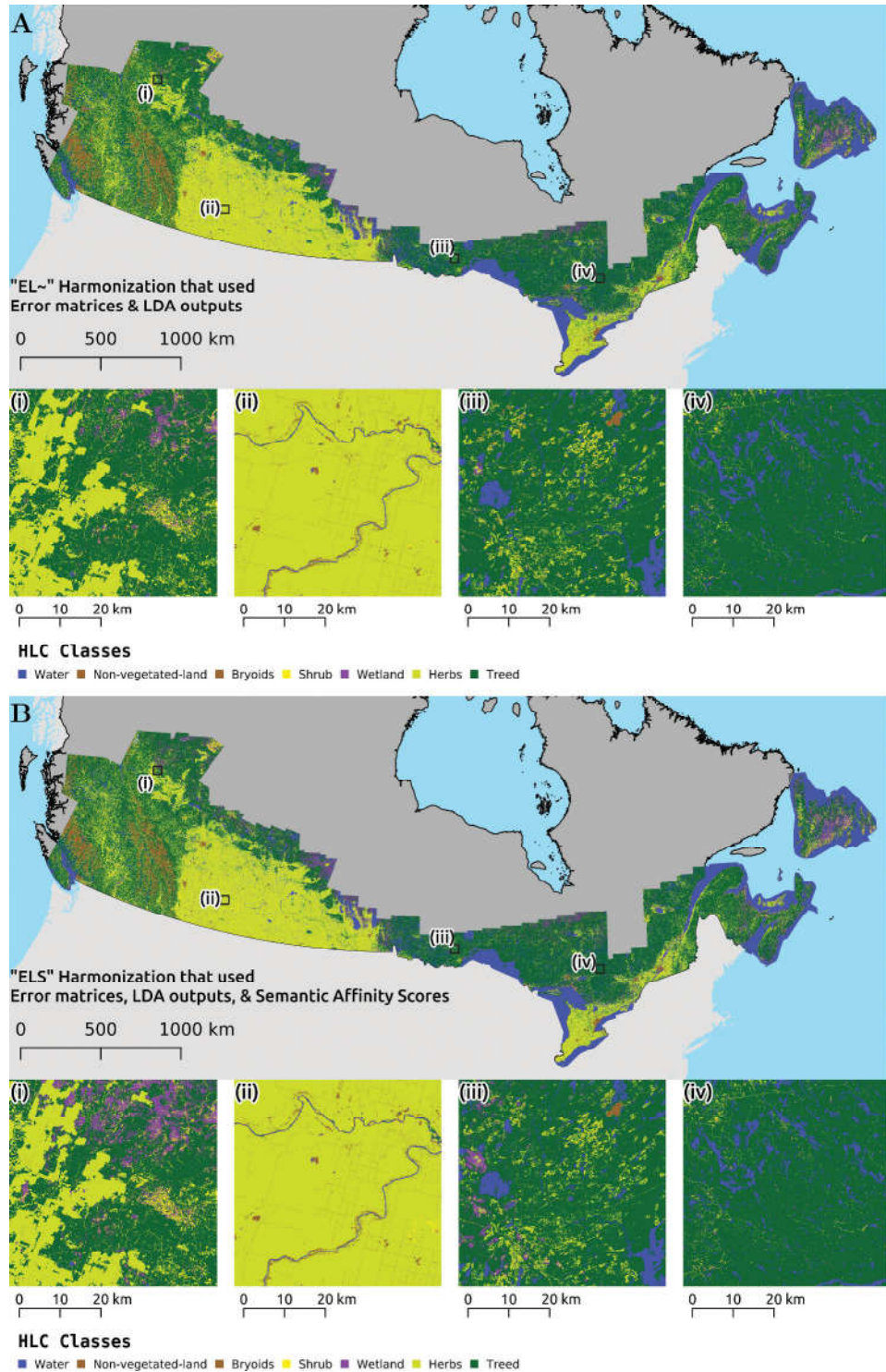


Figure 4. Harmonized maps by the approaches (A) 'EL~2H', using error matrices and LDA outputs for harmonization; and (B) 'ELS2H', using error matrices, LDA outputs, and semantic affinity scores for harmonization.

Table 7. For the area of disagreement, the change in accuracy in the harmonized output maps as a function of not using an error matrix, not using an LDA model, and not using semantic affinity scores. Codes for harmonization scenarios are fully described in Table 1.

		Change in accuracy due to:					
Harmonized Land Cover Class		No Error Matrix			No LDA Model		No Semantic Affinity Score
		$\sim\sim S - E \sim S$	$\sim L \sim - EL \sim$	$\sim LS - ELS$	$\sim\sim S - \sim LS$	$E \sim S - ELS$	$\sim L \sim - \sim LS$
Producer's	Overall	-.199	-.037	-.047	-.097	.054	.029
	Water	-.048	-.256	-.182	.105	-.029	.001
	Non-veg.	.152	-.073	.054	.143	.045	-.056
	Shrub	.527	-.306	-.357	.748	-.136	.022
	Wetland	-.062	.110	.195	-.380	-.122	-.265
	Herbs	-.107	-.026	.012	-.037	.081	.037
	Treed	-.532	.023	-.059	-.362	.111	.105
User's	Water	.050	.286	-.793	.750	-.093	1.000
	Non-veg.	.131	-.095	.151	.115	-.167	-.016
	Shrub	-.128	-.162	-.181	.113	.059	.015
	Wetland	-.020	-.006	-.035	.219	.204	-.046
	Herbs	.008	-.024	-.022	.028	-.001	.008
	Treed	-.121	-.053	-.009	-.043	.069	.008

Discussion

The training of the LDA model uses the co-occurrence of classes between source maps that result either from semantic similarities in class definitions or from class confusion due to classification errors. The LDA model thereby provides similar information content to that of semantic affinity scores and error matrices for land cover harmonization. The difference however is that both error matrices and semantic affinity scores need to be known *before* harmonization as *a priori* information, whereas the LDA-based information is estimated from, and driven by, the source map data directly. As demonstrated in this study, although the LDA outputs are not complete substitutes for error matrices or semantic affinity scores, they can recover sufficient information on class affinities and class confusions to support harmonization efforts when error matrices or semantic affinity scores are unavailable. Furthermore, separate error matrices from source maps characterize confusions between classes *within* each map and may not address class confusion *between* maps, such as the source map error matrices used in this study. However, the LDA model does use co-occurrence of classes between source maps, which helps to address class confusions between maps.

LDA outputs (latent topic probabilities and conditional probabilities of SLC classes for topics) are derived from the statistical modeling of the co-occurrence of SLC classes in source maps. The resulting implicit information on class confusions from LDA outputs complement information on confusion errors of source maps from error matrices that are of poor quality or unavailable (Table 7). Similarly, the LDA provides implicit but data-driven information on semantic affinity. Without using semantic affinity scores, as in 'EL~2H', the LDA statistically connected HLC classes to topics by assigning topics to HLC classes based on combinations of agreed classes (Section 3.4.6). Meanwhile, topics were connected to SLC classes quantitatively by conditional probabilities of SLC classes for given topics. With latent topics as the bridge, we indirectly constructed the association

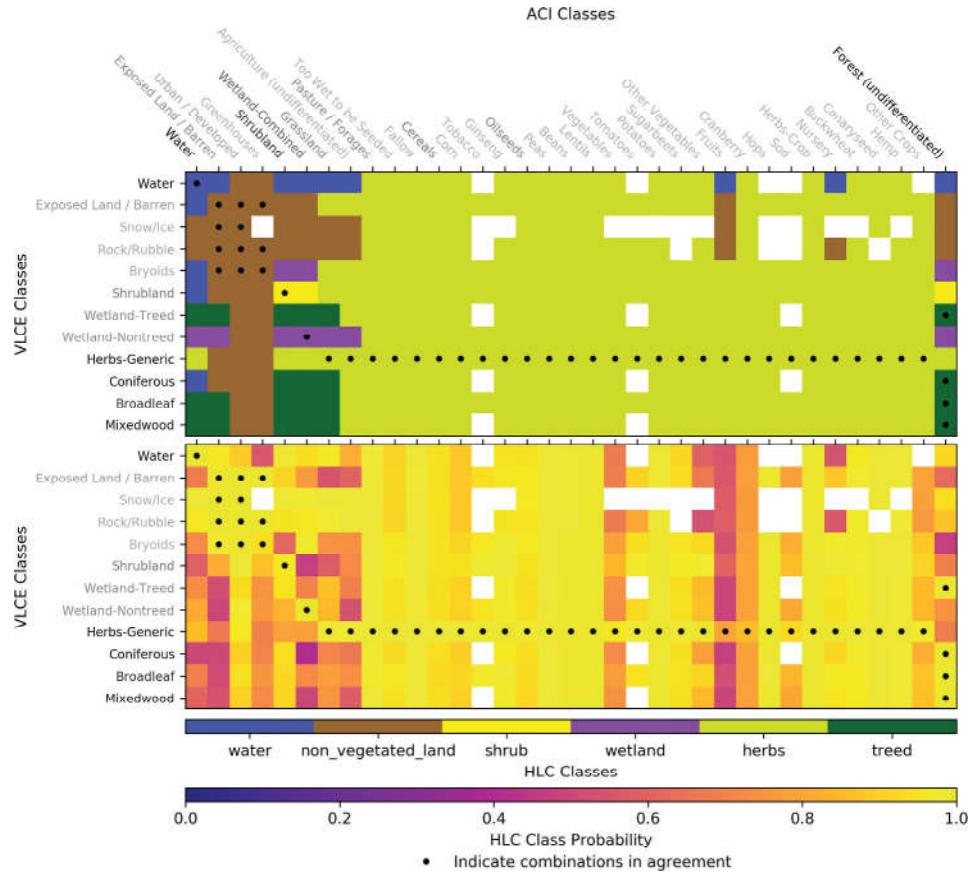


Figure 5. Harmonization by the approach ‘EL~2H’, using error matrices and LDA outputs (Table 1). HLC labels and their class probabilities of all the combinations of source map classes. Blanks indicate no such combination. Darker fonts of class names mean higher frequencies.

between the HLC and SLC classes, i.e. information similar to the rank-based semantic affinity scores between the HLC and SLC classes.

In order to enable LDA model training over map tiles (Section 3.4.4), we assumed that the spatial variations of a landscape within a pixel (30-m spatial resolution) and within an objectively sized tile (300 × 300 pixels) are statistically comparable. If source maps have lower, higher, or different spatial resolutions, further studies would be needed to determine the validity of the assumption of comparable spatial variations of land cover between pixels and tiles. For example, analyzing the sensitivity of harmonization to tile sizes may provide insights. The tile size for training was determined by minimizing the perplexity over test pixels extracted from a simple random sample over the entire map. Some rare classes or class combinations of input maps may be left out by such a sampling protocol, pointing to future opportunities for investigation of stratified random sampling over different class combinations of input maps.

Our results highlight the importance of error matrices in land cover harmonization. This finding concurs with several recent studies that used source map error matrices to achieve

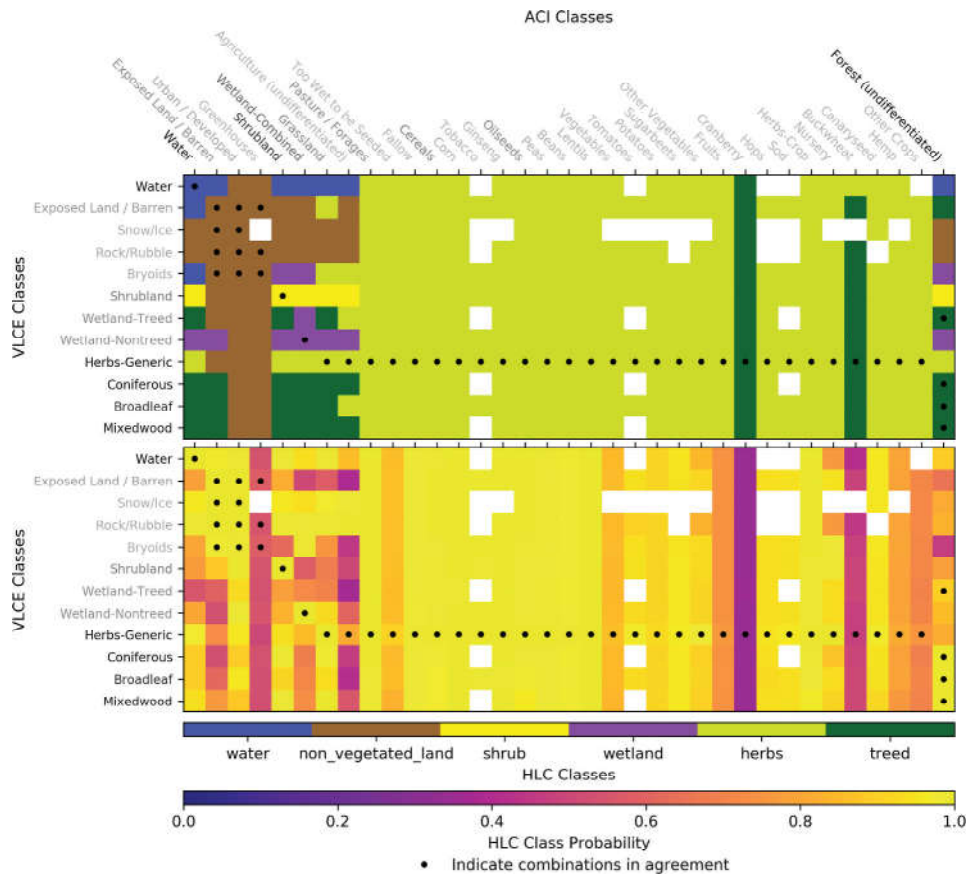


Figure 6. Harmonization by the approach 'ELS2H', using error matrices, LDA outputs, and semantic affinity scores (Table 1). HLC labels and their class probabilities of all the combinations of source map classes. Blanks indicate no such combination. Darker fonts of class names mean higher frequencies.

satisfactory harmonization results (Pérez-Hoyos *et al.* 2012b, Tuanmu and Jetz 2014, See *et al.* 2015, Tsendbazar *et al.* 2015). While our results demonstrate the potential of LDA, our implementation is a single case study, albeit one over a very large area with a range of different environments. In addition, we used error matrices that were spatially global (i.e. representing class accuracies of an entire map as a whole). However, different classes likely have different classification errors over different parts of a map, especially for large-area maps such as those used in this study. Local measures of class accuracy may further improve harmonization. Some recent studies have explored the use of local accuracy measures in harmonization (Schepaschenko *et al.* 2015, See *et al.* 2015, Tsendbazar *et al.* 2015, 2017); however, these local accuracy measures are not class-specific. Many current land cover products provide both classification maps and spatially-explicit class probabilities (sometimes also referred to as class membership, class likelihood, or confidence measures) (Blanco *et al.* 2013, Latifovic *et al.* 2017, Hermosilla *et al.* 2018, Buchhorn *et al.* 2019, Sulla-Menashe *et al.* 2019). These full vectors of class probabilities at the pixel level can be explored with map-level accuracy measures of source maps for harmonization. However, the characterizations of class probabilities in different land cover products are

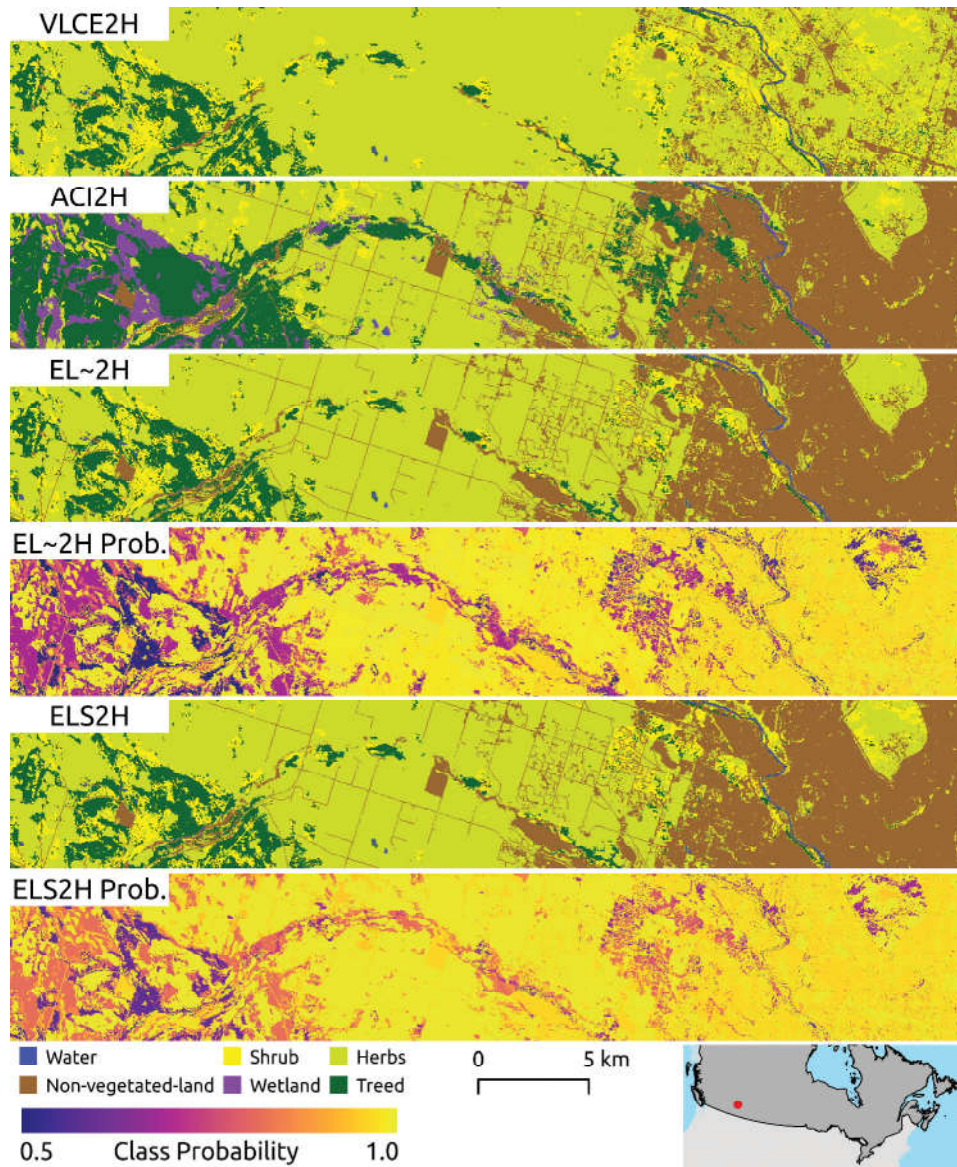


Figure 7. Details of benchmark maps (VLCE2H and ACI2H), harmonized maps (EL~2H and ELS2H), and maps of HLC class probabilities over an example region west of Calgary, Alberta, represented by the red dot in the lower-right map.

usually not similar but rather product specific. Thus, using class probabilities from different products together is like using digital numbers from different sensors, suggesting that some cross calibration of class probabilities between land cover products may be necessary before using them together for harmonization.

Although our case study harmonized two source maps, the LDA-based method also enables the harmonization of three or more maps, a likely case given the increasing availability of land cover maps. More class labels per pixel from more maps provide more

information on pixel-by-pixel co-occurrence of classes among source maps, which helps the LDA model training and the estimation of topic probabilities per pixel. The processing cost of training and running the LDA model will increase, however, given current availability and capacity of high-performance computing facilities we do not envision major computational challenges.

Conclusions

In this study, we introduced the LDA model to the harmonization of two national land cover products over southern Canada respectively focused on forest (VLCE) and agricultural (ACI) environments. We tested multiple harmonization approaches, using the LDA model alone and in combination with more commonly used information sources for land cover harmonization, specifically error matrices and/or semantic affinity scores. The evaluation and comparison of all the approaches revealed both the importance of error matrices and the unique benefits of the LDA model for harmonization. For the harmonization of the source maps in this study, the best performance came from the approach that combined the use of error matrices, LDA outputs, and semantic affinity scores. The harmonized maps generated using this approach combined the strengths of the input source maps while also reducing their weaknesses, as indicated by the achieved overall and user's and producer's accuracies for all the HLC classes. Our results demonstrated that the LDA model boosts the quality of harmonization results and strengthens the robustness of the accuracies of derived harmonized maps, making them less sensitive to the lack of error matrices and/or semantic affinity scores. As more overlapping land cover products from various mapping agencies are generated, the harmonization approaches described in this study provide a framework to help users to take greater advantage of existing thematic maps in a transparent, systematic, and automated manner, to generate high quality harmonized land cover products able to support diverse application-specific objectives in land-related research and management in a cost-effective way.

Acknowledgments

This research was undertaken as part of the "Earth Observation to Inform Canada's Climate Change Agenda (EO3C)" project jointly funded by the Canadian Space Agency (CSA), Government Related Initiatives Program (GRIP), and the Canadian Forest Service (CFS) of Natural Resources Canada. This research was enabled in part by support provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computeCanada.ca). Comber is supported by the Natural Environment Research Council, UK (NE/S009124/1). We thank Dave Hill for his assistance in collecting the validation data, and the Editors and three anonymous reviewers for their constructive review comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Canadian Space Agency [Government Related Initiatives Program (GRIP): ""]; Natural Environment Research Council [NE/S009124/1].

Notes on contributors

Zhan Li is currently a postdoctoral researcher at the German Research Centre for Geosciences, Potsdam, Germany. He has a PhD degree in Geography from Boston University, Boston, USA. Dr. Li previously worked at School for the Environment of University of Massachusetts Boston, Boston, USA, and Pacific Forestry Centre of Canadian Forest Service, Victoria, Canada. His current research interests focus on observing and understanding two aspects of the Earth system, (1) the structure and functioning of terrestrial ecosystems, primarily forests and wetlands; (2) land cover and land use changes and their impacts to and implications for carbon cycling and climate change. He specializes in remote sensing of land at a range of spatial scales by both passive and active sensors.

Joanne White is a Research Scientist with the Canadian Forest Service, Natural Resources Canada. Dr. White received her Doctor of Science degree from the Department of Forest Sciences at the University of Helsinki. Her research interests focus on the use of remotely sensed data, particularly lidar and optical data, for forest inventory and monitoring applications. Dr. White is also an Adjunct Professor in the Faculty of Forestry at the University of British Columbia.

Michael Wulder is Senior Research Scientist with the Canadian Forest Service, Natural Resources Canada. He uses remotely sensed and spatial data to study and monitor forests across Canada, over a range of scales, contributing to national and international programs. He received his PhD from the University of Waterloo and is an Adjunct Professor in the Department of Geography at the University of Victoria and the Department of Forest Resources Management of the University of British Columbia. Dr. Wulder is a member of the USGS/NASA Landsat Science Team (since 2006) and played a lead role in initiating the SilviLaser series of international conferences focused on laser remote sensing of forest environments, co-chairing the 2002 and 2012 editions.

Txomin Hermosilla is a Research Scientist with the Canadian Forest Service, Natural Resources Canada. He received his PhD in Geodesy, Cartography and Geographical Information Systems at the Polytechnic University of Valencia (Universitat Politècnica de València), Spain. Dr. Hermosilla's research interests include the analysis of Landsat time-series data and the development of mass-data processing algorithms and methods to generate detailed datasets that convey forest dynamics.

Andrew Davidson is Manager of Earth Observation in Agriculture Canada's Centre for AgroClimate, Geomatics and Earth Observation (ACGEO). ACGEO is responsible for the operationalization of AAFC's geospatial research. He received his PhD from the Department of Geography at the University of Toronto, and is an Adjunct Research Professor in the Department of Geography and Environmental Studies at Carleton University, where he has taught and supervised students since 2003. Dr. Davidson is also an Affiliated Researcher with the Carleton University Global Water Institute. In 2013, he was elected to the College of Fellows of the Royal Canadian Geographic Society.

Alexis Comber is Professor of Spatial Data Analytics at the University of Leeds, having previously held the chair in Geographical Information Science at the University of Leicester. He completed his PhD in Computer Science at the Macaulay Institute, Aberdeen (now the James Hutton Institute) and the University of Aberdeen developing expert systems for land cover monitoring. Dr. Comber's research interests in spatial data and analysis span many different application areas including land cover / land use, demographics, public health, agriculture, bio-energy and accessibility, all of which require multi-disciplinary approaches.

ORCID

Zhan Li  <http://orcid.org/0000-0001-6307-5200>

Joanne C. White  <http://orcid.org/0000-0003-4674-0373>

Michael A. Wulder  <http://orcid.org/0000-0002-6942-1896>

Txomin Hermosilla  <http://orcid.org/0000-0002-5445-0360>

Andrew M. Davidson  <http://orcid.org/0000-0003-3784-682X>

Alexis J. Comber  <http://orcid.org/0000-0002-3652-7846>

Code availability

The code that supports the findings of this study is available from figshare.com with the identifier: <https://doi.org/10.6084/m9.figshare.11412999>

Data availability statement

The land cover data used in this study are openly available from the National Forest Information System and the Federal Geospatial Platform:

2015 VLCE: https://opendata.nfis.org/downloads/forest_change/CA_forest_VLCE_2015.zip

2015 ACI: <https://open.canada.ca/data/en/dataset/ba2645d5-4458-414d-b196-6303ac06c1c9>

References

- Abercrombie, S.P. and Friedl, M.A., 2016. Improving the consistency of multitemporal land cover maps using a hidden Markov model. *IEEE Transactions on Geoscience and Remote Sensing*, 54, 703–713. doi:10.1109/TGRS.2015.2463689
- Agriculture and Agri-Food Canada, 2018. *Annual crop inventory [WWW document]*. Available from: <https://open.canada.ca/data/en/dataset/ba2645d5-4458-414d-b196-6303ac06c1c9> [Accessed 15 Nov 2018].
- Ahlqvist, O., 2005. Using uncertain conceptual spaces to translate between land cover categories. *International Journal of Geographical Information Science*, 19, 831–857. doi:10.1080/13658810500106729
- Blanco, P.D., et al., 2013. A land cover map of Latin America and the Caribbean in the framework of the SERENA project. *Remote Sensing of Environment*, 132, 13–31. doi:10.1016/j.rse.2012.12.025
- Blei, D.M., 2012. Probabilistic topic models. *Communications of the ACM*, 55, 77. doi:10.1145/2133806.2133826
- Blei, D.M., 2019. *Probabilistic topic models and user behavior [WWW document]*. Available from: http://www.cs.columbia.edu/~blei/talks/Blei_User_Behavior.pdf [Accessed 13 Sept 2019].
- Blei, D.M., Ng, A.Y., and Jordan, M.I., 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Buchhorn, M., et al., 2019. *Copernicus global land service product user manual: moderate dynamic land cover 100m version 2*. Available from: <https://land.copernicus.eu/global/products/lc>
- Cochran, W.G. 1977. *Sampling techniques*. 3rd ed. New York: John Wiley & Sons.
- Comber, A., Fisher, P., and Wadsworth, R., 2004a. Assessment of a semantic statistical approach to detecting land cover change using inconsistent data sets. *Photogrammetric Engineering & Remote Sensing*, 70, 931–938. doi:10.14358/PERS.70.8.931
- Comber, A.J., Fisher, P., and Wadsworth, R., 2004b. Integrating land-cover data with different ontologies: identifying change from inconsistency. *International Journal of Geographical Information Science*, 18, 691–708. doi:10.1080/13658810410001705316
- Comber, A.J., Fisher, P., and Wadsworth, R., 2018. Text mining analysis of land cover semantic overlap. In: O. Ahlqvist, D. Varanka, S. Fritz, K. Janowicz, ed. *Land use and land cover semantics*. Boca Raton: CRC Press, 208–227.
- Davidson, A.M., et al., 2017. Detailed crop mapping using remote sensing data (crop data layers). In: J. Delincé, ed. *Handbook on remote sensing for agricultural statistics*. Chapter 4. Rome: Handbook of the Global Strategy to improve Agricultural and Rural Statistics (GSARS), 91–130.
- Feng, C.C. and Flewelling, D.M., 2004. Assessment of semantic similarity between land use/land cover classification systems. *Computers, Environment and Urban Systems*, 28, 229–246. doi:10.1016/S0198-9715(03)00020-6

- Franklin, S.E. and Wulder, M.A., 2002. Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography: Earth and Environment*, 26, 173–205. doi:[10.1191/0309133302pp332ra](https://doi.org/10.1191/0309133302pp332ra)
- Fritz, S., et al., 2011. Highlighting continued uncertainty in global land cover maps for the user community. *Environmental Research Letters*, 6, 044005. doi:[10.1088/1748-9326/6/4/044005](https://doi.org/10.1088/1748-9326/6/4/044005)
- Fritz, S. and See, L., 2008. Identifying and quantifying uncertainty and spatial disagreement in the comparison of Global Land Cover for different applications. *Global Change Biology*, 14, 1057–1075. doi:[10.1111/j.1365-2486.2007.01519.x](https://doi.org/10.1111/j.1365-2486.2007.01519.x)
- Giri, C., Zhu, Z., and Reed, B., 2005. A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. *Remote Sensing of Environment*, 94, 123–132. doi:[10.1016/j.rse.2004.09.005](https://doi.org/10.1016/j.rse.2004.09.005)
- Griffiths, T.L. and Steyvers, M., 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101, 5228–5235. doi:[10.1073/pnas.0307752101](https://doi.org/10.1073/pnas.0307752101)
- Hermosilla, T., et al., 2015. An integrated Landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. *Remote Sensing of Environment*, 158, 220–234. doi:[10.1016/j.rse.2014.11.005](https://doi.org/10.1016/j.rse.2014.11.005)
- Hermosilla, T., et al., 2018. Disturbance-informed annual land cover classification maps of Canada's forested ecosystems for a 29-year landsat time series. *Canadian Journal of Remote Sensing*, 44, 1–21. doi:[10.1080/07038992.2018.1437719](https://doi.org/10.1080/07038992.2018.1437719)
- Herold, M., et al., 2006. A joint initiative for harmonization and validation of land cover datasets. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1719–1727. doi:[10.1109/TGRS.2006.871219](https://doi.org/10.1109/TGRS.2006.871219)
- Herold, M., et al., 2008. Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, 112, 2538–2556. doi:[10.1016/j.rse.2007.11.013](https://doi.org/10.1016/j.rse.2007.11.013)
- Iwao, K., et al., 2011. Creation of new global land cover map with map integration. *Journal of Geographical Information Systems*, 03, 160–165. doi:[10.4236/jgis.2011.32013](https://doi.org/10.4236/jgis.2011.32013)
- Jepsen, M.R. and Levin, G., 2013. Semantically based reclassification of Danish land-use and land-cover information. *International Journal of Geographical Information Science*, 27, 2375–2390. doi:[10.1080/13658816.2013.803555](https://doi.org/10.1080/13658816.2013.803555)
- Jung, M., et al., 2006. Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sensing of Environment*, 101, 534–553. doi:[10.1016/j.rse.2006.01.020](https://doi.org/10.1016/j.rse.2006.01.020)
- Kaptué Tchuenté, A.T., Roujean, J.-L., and De Jong, S.M., 2011. Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. *International Journal of Applied Earth Observation and Geoinformation*, 13, 207–219. doi:[10.1016/j.jag.2010.11.005](https://doi.org/10.1016/j.jag.2010.11.005)
- Kavouras, M. and Kokla, M., 2002. A method for the formalization and integration of geographical categorizations. *International Journal of Geographical Information Science*, 16, 439–453. doi:[10.1080/13658810210129120](https://doi.org/10.1080/13658810210129120)
- Kinoshita, T., Iwao, K., and Yamagata, Y., 2014. Creation of a global land cover and a probability map through a new map integration method. *International Journal of Applied Earth Observation and Geoinformation*, 28, 70–77. doi:[10.1016/j.jag.2013.10.006](https://doi.org/10.1016/j.jag.2013.10.006)
- Latifovic, R., Pouliot, D., and Olthof, I., 2017. Circa 2010 land cover of Canada: local optimization methodology and product development. *Remote Sensing*, 9, 1098. doi:[10.3390/rs9111098](https://doi.org/10.3390/rs9111098)
- Lesiv, M., et al., 2016. Comparison of data fusion methods using crowdsourced data in creating a hybrid forest cover map. *Remote Sensing*, 8 (3), 261. doi:[10.3390/rs8030261](https://doi.org/10.3390/rs8030261)
- Matthews, E., 1983. Global vegetation and land use: new high-resolution data bases for climate studies. *Journal of Climate and Applied Meteorology*, 22, 474–487. doi:[10.1175/1520-0450\(1983\)022<0474:GVALUN>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<0474:GVALUN>2.0.CO;2)
- McCallum, I., et al., 2006. A spatial comparison of four satellite derived 1 km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation*, 8, 246–255. doi:[10.1016/j.jag.2005.12.002](https://doi.org/10.1016/j.jag.2005.12.002)
- Olofsson, P., et al., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42–57. doi:[10.1016/j.rse.2014.02.015](https://doi.org/10.1016/j.rse.2014.02.015)
- Pedregosa, F., et al., 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

- Pérez-Hoyos, A., et al., 2017. Comparison of global land cover datasets for cropland monitoring. *Remote Sensing*, 9, 1118. doi:[10.3390/rs9111118](https://doi.org/10.3390/rs9111118)
- Pérez-Hoyos, A., García-Haro, F.J., and San-Miguel-Ayanz, J., 2012a. Conventional and fuzzy comparisons of large scale land cover products: application to CORINE, GLC2000, MODIS and GlobCover in Europe. *ISPRS Journal of Photogrammetry and Remote Sensing*, 74, 185–201. doi:[10.1016/j.isprsjprs.2012.09.006](https://doi.org/10.1016/j.isprsjprs.2012.09.006)
- Pérez-Hoyos, A., García-Haro, F.J., and San-Miguel-Ayanz, J., 2012b. A methodology to generate a synergetic land-cover map by fusion of different land-cover products. . *International Journal of Applied Earth Observation and Geoinformation*, 19, 72–87. doi:[10.1016/j.jag.2012.04.011](https://doi.org/10.1016/j.jag.2012.04.011)
- Pflugmacher, D., et al., 2011. Comparison and assessment of coarse resolution land cover maps for Northern Eurasia. *Remote Sensing of Environment*, 115, 3539–3553. doi:[10.1016/j.rse.2011.08.016](https://doi.org/10.1016/j.rse.2011.08.016)
- Running, S.W., 2008. CLIMATE CHANGE: ecosystem disturbance, carbon, and climate. *Science*, 80, 321, 652–653. doi:[10.1126/science.1159607](https://doi.org/10.1126/science.1159607)
- Schepaschenko, D., et al., 2015. Development of a global hybrid forest mask through the synergy of remote sensing, crowdsourcing and FAO statistics. *Remote Sensing of Environment*, 162, 208–220. doi:[10.1016/j.rse.2015.02.011](https://doi.org/10.1016/j.rse.2015.02.011)
- See, L., et al., 2015. Building a hybrid land cover map with crowdsourcing and geographically weighted regression. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 48–56. doi:[10.1016/j.isprsjprs.2014.06.016](https://doi.org/10.1016/j.isprsjprs.2014.06.016)
- Sellers, P.J., et al., 1997. Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science*, 80 (275), 502–509. doi:[10.1126/science.275.5299.502](https://doi.org/10.1126/science.275.5299.502)
- Stehman, S.V. and Czaplewski, R.L., 2003. Introduction to special issue on map accuracy. *Environmental and Ecological Statistics*, 10, 301–308. doi:[10.1023/A:1025138423071](https://doi.org/10.1023/A:1025138423071)
- Sulla-Menashe, D., et al., 2019. Hierarchical mapping of annual global land cover 2001 to present: the MODIS Collection 6 Land Cover product. *Remote Sensing of Environment*, 222, 183–194. doi:[10.1016/j.rse.2018.12.013](https://doi.org/10.1016/j.rse.2018.12.013)
- Townshend, J.R.G., 1992. Land cover. *International Journal of Remote Sensing*, 13, 1319–1328. doi:[10.1080/01431169208904193](https://doi.org/10.1080/01431169208904193)
- Tsendbazar, N.-E., et al., 2015. Spatial accuracy assessment and integration of global land cover datasets. *Remote Sensing*, 7, 15804–15821. doi:[10.3390/rs71215804](https://doi.org/10.3390/rs71215804)
- Tsendbazar, N.-E., de Bruin, S., and Herold, M., 2017. Integrating global land cover datasets for deriving user-specific maps. *International Journal of Digital Earth*, 10, 219–237. doi:[10.1080/17538947.2016.1217942](https://doi.org/10.1080/17538947.2016.1217942)
- Tuanmu, M.N. and Jetz, W., 2014. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography*, 23, 1031–1045. doi:[10.1111/geb.12182](https://doi.org/10.1111/geb.12182)
- Vancutsem, C., et al., 2013. Harmonizing and combining existing land cover/land use datasets for cropland area monitoring at the African continental scale. *Remote Sensing*, 5, 19–41. doi:[10.3390/rs5010019](https://doi.org/10.3390/rs5010019)
- Wadsworth, R.A., Comber, A.J., and Fisher, P.F., 2006. Expert knowledge and embedded knowledge: or why long rambling class descriptions are useful, in: *Progress in Spatial Data Handling - 12th International Symposium on Spatial Data Handling*, 12–14 July 2006. Vienna, Austria: SDH. doi:[10.1007/3-540-35589-8_13](https://doi.org/10.1007/3-540-35589-8_13)
- White, J.C., et al., 2014. Pixel-based image compositing for large-area dense time series applications and science. *Canadian Journal of Remote Sensing*, 40, 192–212. doi:[10.1080/07038992.2014.945827](https://doi.org/10.1080/07038992.2014.945827)
- Wulder, M.A., et al., 2008. Monitoring Canada's forests. Part 1: completion of the EOSD land cover project. *Can. Journal of Remote Sensing*, 34, 549–562. doi:[10.5589/m08-066](https://doi.org/10.5589/m08-066)
- Wulder, M.A., et al., 2018. Land cover 2.0. Int. *International Journal of Remote Sensing*, 39, 4254–4284. doi:[10.1080/01431161.2018.1452075](https://doi.org/10.1080/01431161.2018.1452075)