CrossMark

Image interpretation-guided supervised classification using nested segmentation

A.V. Egorov [a], M.C. Hansen [b,*], D.P. Roy [a], A. Kommareddy [b], P.V. Potapov [b]

[a] Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD 57007, USA
[b] University of Maryland, College Park, MD, USA

ABSTRACT

We present a new binary (two-class) supervised non-parametric classification approach that is based on iterative partitioning of multidimensional feature space into variably-sized and nested hyper-cubes (partitions). The proposed method contains elements of active learning and includes classifier to analyst queries. The spectral transition zone between two thematic classes (i.e., where training labels of different classes overlap in feature space) is targeted through iterative training derivation. Three partition categories are defined: *pure*, *indivisible* and *unlabeled*. *Pure* partitions contain training labels from only one class, *indivisible* partitions contain training data from different classes, and *unlabeled* partitions do not contain training data. A minimum spectral tolerance threshold defines the smallest partition volume to avoid over-fitting. In this way the transition zones between class distributions are minimized, thereby maximizing both the spectral volume of *pure* partitions in the feature space and the number of *pure* pixels in the classified image. The classification results are displayed to show each classified pixel's partition category (*pure*, *unlabeled* and *indivisible*). Mapping pixels belonging to *unlabeled* partitions serves as a query from the classifier to the analyst, targeting spectral regions absent of training data. The classification process is repeated until significant improvement of the classification is no longer realized or when no classification errors and *unlabeled* pixels are left. Variably-sized partitions lead to intensive training data derivation in the spectral transition zones between the target classes. The methodology is demonstrated for surface water and permanent snow and ice classifications using 30 m conterminous United States Landsat 7 Enhanced Thematic Mapper Plus (ETM+) data time series from 2006 to 2010. The surface water result was compared with Shuttle Radar Topography Mission (SRTM) water body and National Land Cover Database (NLCD) open water classes with an overall agreement greater than 99% and Kappa coefficient greater than 0.9 in both of cases. In addition, the surface water result was compared with a classification generated using the same input data and a standard bagged Classification and Regression Tree (CART) classifier. The nested segmentation and CART-generated products had an overall agreement of 99.9 and Kappa coefficient of 0.99.

1. Introduction

Classification is regarded as a fundamental process in remote sensing used to relate pixel values to land cover or sometimes land use classes that are present at the corresponding location on the Earth's surface (Mather, 2004). Conventionally, pixel class assignment is determined by the spectral properties (signatures) of a given class or theme. Each spectral feature, for example red, near-infrared or shortwave infrared reflectance, is taken as an explanatory or independent variable. The theoretical n-dimensional space where n axes correspond to n raster bands in multispectral imagery, or n band transformations extracted from single images or time series, is often termed the feature space. Classifiers assign labels to pixels based on partitioning of feature space values using either unsupervised or training-based supervised methods.

Supervised classification methods have a long history since the development of techniques such as linear discriminant analysis (LDA) to classify two or more sub-populations (Fisher, 1936). Numerous classification algorithms have been developed and those applied to remotely sensed data include: k-nearest neighbor (kNN) (Fix & Hodges, 1951), multilayer perceptron (MLP) (Rosenblatt, 1957, 1958), maximum likelihood (ML) (Savage, 1976), Kohonen's self organized map (SOM) (Kohonen, 1982; Kohonen & Honkela, 2007), classification and regression trees (CART) (Breiman, Friedman, Olshen, & Stone, 1984), support vector machine (SVM) (Cortes & Vapnik, 1995), and random forests (RF) (Breiman, 2001). In supervised classification methods, training data of accurately labeled examples are taken as the dependent variable and associated to a set of independent variables. For land cover mapping using earth observation imagery, training data may be gathered on the basis of image interpretation, ground measurements or any other trusted source of information. In general, collecting training

* Corresponding author.

data requires considerable time and effort. Supervised classification approaches are dependent on the experience of the remote sensing analyst in collecting training data and on the quality of the imagery. Supervised methods require *a priori* knowledge of the feature of investigation (e.g., the land cover type) in order to derive appropriate training data. Generating a training data set that accounts for all relevant spectral heterogeneity within and between classes is challenging and no systematic approach exists for training data collection. For example, training data selected by an analyst in the field may not be sufficiently representative of the conditions encountered in the image. Quality training data are required to achieve accurate supervised classification results.

Semi-automatic training set derivation has the goal of producing a parsimonious but sufficient set of training labels for supervised classification. Usually the acquisition of labeled data is difficult, time-consuming, or expensive to obtain. For these reasons a training set should be kept small while ensuring adequate classification performance. Several studies have shown however that classification accuracy increases with training set size (Lippitt, Rogan, Li, Eastman, & Jones, 2008; Rogan et al., 2008; Yan & Roy, 2015), although the optimal training size and distribution are usually unknown (Arora & Foody, 1997; Foody & Mathur, 2004b; Foody, McCulloch, & Yates, 1995; Pal & Mather, 2003; Zhuang, Engel, Lozanogarcia, Fernandez, & Johannsen, 1994). Many studies have emphasized the positioning of training data within the feature space, particularly the importance of collecting both pure (only one class in the pixel) and mixed pixel (more than one class in the pixel) training data. For example, Foody and Mathur (2004a,b, 2006) showed that the acquisition of training samples near feature space class boundaries may help reduce the training data set size without a loss of SVM classification accuracy. Similarly, Yu and Chi (2008) showed that a small training data set collected along class spectral boundaries provided comparable SVM classification accuracy to using training data consisting of a large number of pure pixels. Tuia, Pacifici, Kanevski, and Emery (2009) likewise employed a SVM and active learning to generate training data in classifying a series of single images. Other studies have shown similar results using mixed pixel training with aNN (Bernard, Wilkinson, & Kanellopoulos, 1997; Foody, 1999) and CART (Hansen, 2012) classifiers. Thus, a training set should be kept small, when training data collection is expensive, and should include both pure and mixed training data with particular emphasis on training data collection at the feature space class boundaries.

Semi-automatic training set derivation has been referred to as "active learning" in the machine learning literature and as "query learning" or "optimal experimental design" in the statistics literature (Settles, 2009). Active learning focuses on the interaction between the analyst (or some other information source) and the classifier. The model returns to the analyst the pixels whose classification outcome is the most uncertain. After accurate labeling by the analyst, pixels are added to the training set in order to reinforce the model. In this way, the model is optimized on well-chosen difficult examples, maximizing its generalization capabilities (Tuia, Volpi, Copa, Kanevski, & Munoz-Mari, 2011). Semi-automatic learning can be of great practical value in many real-word problems where unlabeled data are abundant or easily obtained, but the acquisition of labeled data is difficult, time-consuming, or expensive to obtain (Lippitt et al., 2008; Settles, 2009). Active learning algorithms have been studied in many real world problems, such as classifying handwritten characters (Lang & Baum, 1992), part-of-speech tagging (Dagan & Engelson, 1995), sensor scheduling (Krishnamurthy, 2002), learning ranking functions for information retrieval (Yu, 2005), word sense disambiguation (Fujii, Tokunaga, Inui, & Tanaka, 1998), text classification (Hoi, Jin, & Lyu, 2006; Lewis & Catlett, 1994; McCallum & Nigam, 1998; Tong & Koller, 2000), information extraction (Settles & Craven, 2008; Thompson, Califf, & Mooney, 1999), video classification and retrieval (Hauptmann, Lin, Yan, Yang, & Chen, 2006; Yan, Yang, & Hauptmann, 2003), speech recognition (Tür et al., 2005), and cancer diagnosis (Liu, 2004). Active learning is also

suitable for remote sensing applications, where the number of pixels among which the search is performed is large and manual definition is redundant and time consuming. However, only a relatively few studies have been dedicated to remote sensing data classification using active learning (e.g. Jackson & Landgrebe, 2001; Jun & Ghosh, 2008; Li, Bioucas-Dias, & Plaza, 2010; Licciardi et al., 2009; Tuia et al., 2009, 2011).

This study builds on previous research by presenting a semi-automatic active learning classification approach called *nested segmentation*. Nested segmentation identifies areas in need of labeling followed by manual assignment by an analyst. The resulting systematic feature space partitioning defines the classification rules, i.e., unlike other active learning classification approaches (Tuia et al., 2009) an extant classification algorithm is not used. The approach is iterated until either a preset classification accuracy is acquired or there are no unlabeled classified pixels. Instead of relying simply on the size of the training data set to produce a quality classification, we focus on two other training set properties, *representativeness* and *concentration*. Training data that sufficiently cover the intra-class spectral variation per land cover type are representative. Training data that are densely located along spectral class boundaries are concentrated. Training data representativeness is achieved by identifying and adding training data in regions of the feature space that lack training samples. Training data concentration is achieved by identifying regions of the feature space where different classes overlap, targeting the addition of training data and recursively sub-dividing the particular spectral region. This allows the analyst's efforts to be focused on deriving training where more intensive sampling is needed. The method provides a new way of iteratively collecting training data for a binary classification that allows an analyst to collect a compact and sufficient training data set.

The nested segmentation approach is designed to be fast in its implementation and appropriate for large area mapping tasks at national to global scales that normally require large training data sets. Mapping at such scales presents a challenge for training data set derivation due to the variety of intra- and interclass spectral variation present. For example, at national scales, surface water can range from clearly identifiable low turbidity lakes to more challenging water bodies, including turbulent coastal surface waters and briny inland lakes of endorheic basins. Land covers such as dark conifer forests or central business districts featuring tall buildings can be confused with open water bodies. The presented method is meant to target all such variations in a rapid, iterative fashion. The methodology is first described and then demonstrated by application to 5 years of 30 m conterminous United States Landsat 7 Enhanced Thematic Mapper Plus (ETM+) Web Enabled Landsat (WELD) data (Roy et al., 2010) to generate open surface water (SW) and permanent snow and ice (SI) classifications. The SW classification is compared quantitatively with water masks from the Shuttle Radar Topography Mission (SRTM) water body data set (Rabus, Eineder, Roth, & Balmer, 2003) and the National Land Cover Database (NLCD2006) open water class (Fry et al., 2011). In addition, the WELD nested segmentation SW classification is compared with a SW classification generated from the same training and Landsat data but using a standard bagged CART classifier. This is followed by a brief discussion of the methodology and implications for future research.

2. Data and pre-processing

2.1. Landsat data

The Landsat satellite series, operated by the U.S. Department of Interior/U.S. Geological Survey (USGS) Landsat project, with satellite development and launches engineered by the National Aeronautics and Space Administration (NASA), represent the longest dedicated land remote sensing data record (Roy, Wulder, et al., 2014). Landsat data provide a balance between requirements for localized moderate spatial resolution studies and global monitoring (Goward, Masek, Williams, Irons, & Thompson, 2001). Free of charge radiometrically

and terrain corrected Landsat data, available through the USGS Center for Earth Resources Observation and Science (EROS) (Woodcock et al., 2008), are the choice of many performing land cover mapping at regional, continental and global scales (Hansen & Loveland, 2012). For example, Landsat data have been used to generate the 21 class 30 m National Land Cover Dataset for the conterminous United States (CONUS) Alaska and Hawaii for 1992, 2001 and 2006 (Fry et al., 2011; Vogelmann et al., 2001). The PRODES Project (Projeto de Monitoramento do Desflorestamento na Amazonia Legal), conducted by Brazil's National Institute for Space Research (INPE), has been using Landsat data to monitor deforestation rates across the Brazilian Amazon annually since 1988 (INPE, 2013). The U.S. Department of Agricultural (USDA) uses Landsat and Landsat-like satellite data to monitor cropping systems domestically and abroad and produces an annual CONUS Cropland Data Layer (CDL) that defines over 100 land cover and crop type classes at 30 m (Johnson & Mueller, 2010).

Weekly CONUS Landsat data provided by the Web-Enabled Landsat Data (WELD) were used for this study (http://e4ftl01.cr.usgs.gov/WELD/). The CONUS WELD Version 1.5 data were generated using every Landsat 7 Enhanced Thematic Mapper Plus (ETM+) Level 1T acquisition with cloud cover ≤80% available from the U.S. Landsat archive (Roy et al., 2010). Version 1.5 WELD data have been used to generate 30 m CONUS annual land cover (Hansen et al., 2011) and 5-year land cover change (Hansen et al., 2014) classifications.

The Version 1.5 WELD weekly products for weeks 16 to 46 (April 15 to November 17) were used to capture the main CONUS growing season, and to avoid weeks that are typically more cloud contaminated at the time of Landsat 7 overpass (Ju & Roy, 2008). Five years of products from 2006 to 2010 were used, providing a total of 155 weeks. Each weekly product contains 14 30 m bands—top of atmosphere (TOA) reflectance for blue (0.45–0.52 μm), green (0.53–0.61 μm), red (0.63–0.69 μm), near-infrared (0.78–0.90 μm), mid-infrared (1.55–1.75 μm and 2.09–2.35 μm), and low and high gain brightness temperature (10.40–12.50 μm), TOA normalized difference vegetation index (NDVI), the date of each acquisition, the per-band radiometric saturation status and two cloud mask values. The CONUS products are defined in 501 tiles of 5000 × 5000 30 m pixels in the Albers equal area projection.

2.2. Classification metrics

Temporal metrics have been shown to be a viable transformation of time-series data to provide feature space variables for land cover and land cover change classification using both coarse resolution (DeFries et al., 1995; Hansen et al., 2008; Reed et al., 1994) and moderate resolution Landsat time-series (Broich et al., 2011; Hansen et al., 2013; Potapov et al., 2012). Metrics are selected to capture seasonal class spectral variations in a way that is robust to missing data and to reduce residual cloud, shadow and atmospheric contamination (Broich et al., 2011; DeFries et al., 1995; Hansen et al., 2011, 2014). In this study, median 5 year metrics, specifically the median value from the 155 weeks at each pixel location, were derived for bands 3 (0.63–0.69 μm), 4 (0.78–0.90 μm), 5 (1.55–1.75 μm), and 7 (2.09–2.35 μm). The blue (0.45–0.52 μm) and green bands (0.53–0.61 μm) were not used due to their sensitivity to atmospheric effects (Roy, Qin, et al., 2014). In this way only four metrics were used. In addition, for post-classification processing purposes, the median Landsat high gain brightness temperature (10.40–12.50 μm) over 155 weeks at each pixel location was also derived. Pixels with no data, due to the scan line corrector issue, and pixels flagged as cloudy, were excluded from metrics generation.

3. Methods

A new supervised active learning classification approach is presented. The method is developed specifically for the classification of two classes and allows an analyst to build a representative and concentrated training data set. The process requires a conventional initial training data set that is sampled from the most obvious and indisputable areas, similar to the approach of Tuia et al. (2009); for example, for the open SW classification initial water training pixels were selected from the centers of deep lakes and rivers with no sediment or weeds and the non-water training pixels were selected from deserts, forests, and bare rocks.

After initial training data collection, an iterative procedure is followed. The feature space is divided automatically into nested variably-sized hyper-cube partitions that have dimensions no smaller than a predefined minimum spectral tolerance threshold. This partitioning results in a set of rules to be applied to the metrics for the image data. Classification results are displayed to show the association of each pixel to the category of the partition to which it belongs. The partition category may be *pure* (all training pixels in the partition belong to the same class), *indivisible* (training pixels in the partition belong to both classes), or *unlabeled* (there are no training pixels in the partition). The analyst refines the training data and the classification process is iteratively repeated. The process can be stopped either when no classification errors and unlabeled pixels are left, or when the desired classification accuracy is reached. The number of iterations of this process is determined by the analyst; after several iterations, the classification quality stabilizes and iteration is no longer pursued. These steps are described below in Sections 3.1–3.3.

3.1. Classification model generation by automated nested feature space partitioning

The feature space is automatically and recursively divided into nested hyper-cube partitions by examination of the training data, as illustrated in Fig. 1. The WELD reflectance data are stored with a 10,000 scaling factor; nominally the reflectance is defined in dimensionless units in the range 0 to 1 and so the data are stored with values from 0 to 10,000. The partitioning algorithm successively splits the feature space into equal halves along each metric; therefore it is most convenient to consider the feature space ranging in value from 0 to 2^{14} (16,384). In this way, the spectral space may be divided into equal integer multiples of two with a minimum partition size of 2^0.

The initial partition (Fig. 1a) is a single hyper-cube defined for the four metric bands (i.e., the 5-year median values of Landsat ETM+ bands 3, 4, 5 and 7 respectively) with feature space side length coordinates from 0 to 16,384. The hyper-cube is then split in half along each metric (for a two dimensional feature space this means splitting the entire feature space into four quarters) (Fig. 1b). The equal splitting procedure is repeated many times, as illustrated in Fig. 1c to e for the *mixed* partitions that contain training pixels of both classes. The recursive procedure is stopped when there are either no *mixed* partitions, or all *mixed* partitions have side lengths equal to the minimum spectral tolerance threshold. *Mixed* partitions that have side lengths equal to the minimum spectral tolerance threshold are termed *indivisible* partitions (dark magenta, Fig. 1f). This recursive process effectively results (e.g., Fig. 1f) in partitions of varying size and with partition boundaries where the two classes in the feature space overlap or are closely adjacent. The partition boundaries define the classification rules. Any pixel falling within a *pure* partition in the feature space is assigned to the corresponding class. Pixels that fall within an *unlabeled* partition are not classified. Pixels that fall within *indivisible* partitions are categorized as *indivisible*. The minimum spectral tolerance threshold is the only *a priori* defined parameter used in the automated nested feature space partitioning process. In this research, different minimum spectral tolerance thresholds were used for the open surface water (SW) and permanent snow and ice (SI) classifications as these classes have different reflectance values in Landsat ETM+ bands 3, 4, 5 and 7. The Landsat-7 ETM+ sensor radiometric calibration uncertainties are estimated as 5% for all the reflective wavelength bands (Markham & Helder, 2012). Consequently, highly reflective surfaces, such as snow and ice, have higher greater absolute reflectance uncertainty than low reflectance
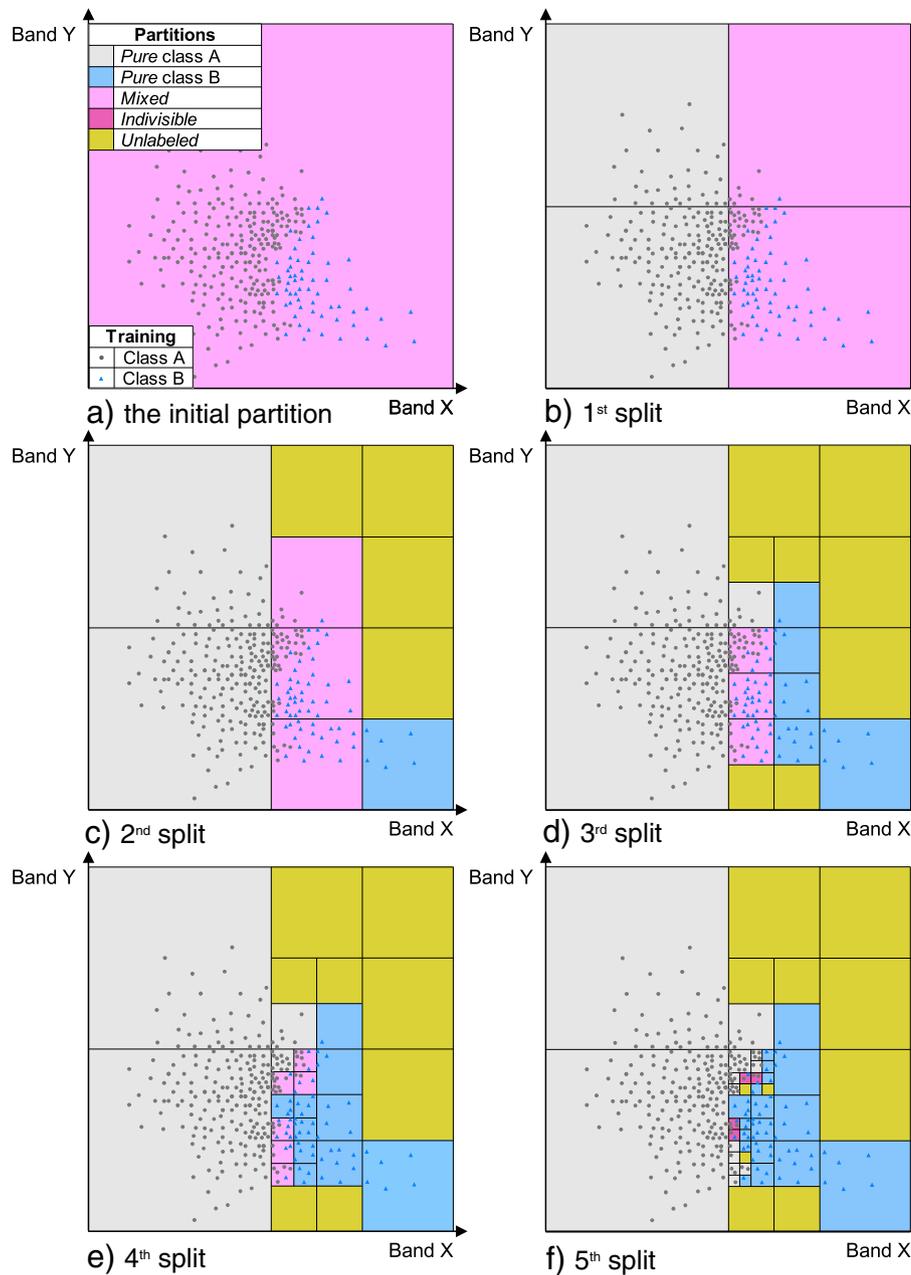
Fig. 1. Feature space illustration of the classification model generated by automated nested partitioning. For illustrative clarity only a two dimensional feature space is shown using synthetic (not real) data. The training data are shown as dots (class A is gray circles and class B is blue triangles) and the partitions are shown as squares.

surfaces such as water. Minimum spectral tolerance thresholds of 32 and 128 (which correspond to 0.0032 and 0.0128 defined in reflectance units with a 10,000 scaling factor) were used for the SW and SI classifications respectively.

3.2. Classification and result inspection

The classification rules defined by the automated nested feature space partitioning (as shown in Fig. 1f) are applied to the metrics for the image data. The resulting classification is displayed with four colors that show the association of each classified pixel to the class and category of the partition in which it fell, i.e., *pure* class A, *pure* class B, *unlabeled*, or *indivisible* (could be either class A or B). The median band 5, median band 4 and median band 3 metrics are also displayed as a false color composite to provide spatial context (Fig. 2); other metrics combinations could also be displayed as desired by the analyst. The analyst

refines the training data by examination of the displayed results (Fig. 2). This is described in Section 3.3.

3.3. Refined training data collection

If the initial training data collection was insufficiently representative, the classification results can be poor and more training data must be added. The analyst adds training data by examination of the classification results (Section 3.2) at pixel locations belonging to *unlabeled* partitions (yellow pixels in Fig. 2 left) and also where the classification is judged visually to be incorrect. For convenience we term these training data collection steps as *gap-filling* and *error-fixing* respectively.

The analyst does not examine the feature space when refining the training data collection. However, it is helpful to consider the partitioning of the feature space before and after new training data are collected to understand the nested segmentation algorithm. This is
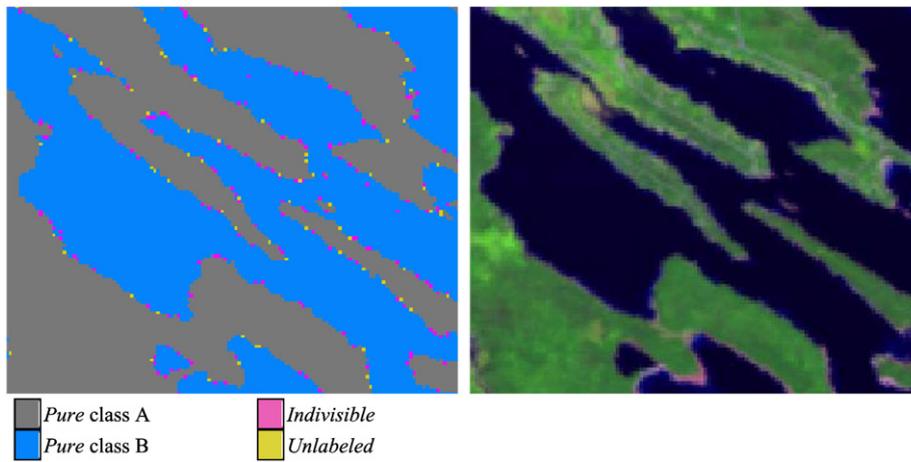
Fig. 2. Left: classification results shown using the coloring scheme illustrated in Fig. 1(f). Right: 5-year median metrics (bands 5, 4, 3 as RGB).

illustrated in Fig. 3 which shows the feature space partitioning before (a) and after (b) new training data (shown as outlined dots) are added. The result of applying the automated nested feature space partitioning after the new training data are added is clearly apparent when comparing Fig. 3(a) and (b). There are more partitions, particularly in the spectral transition zone between the two thematic classes, and there are fewer *unlabeled* partitions. Four illustrative cases annotated in Fig. 3 are described below.

Case 1—*Gap-filling* correction. In the previous classification iteration (Fig. 3a), this partition was *unlabeled* as there were no training pixels within it. Pixels from the classified image falling into this partition were categorized as *unlabeled* in the classified map (e.g., yellow colored pixels in Fig. 2a). Consequently, a new Class A training pixel was added to the partition. After the application of the automated nested feature space partitioning the partition was classified as *pure* class A (Fig. 3b). Consequently, all pixels in the new classified image that fell within this partition were classified as *pure* class A.

Case 2—*gap-filling* correction. In the previous classification iteration (Fig. 3a), this partition was *unlabeled* as there were no training pixels within it. Additional training data resulted in new training pixels of both classes being added to the partition. After automated nested feature space partitioning the partition still contained training data of both classes (Fig. 3b) and was therefore categorized as *indivisible*. Consequently, all pixels in the classified image that fell within this partition were labeled as *indivisible* (i.e. could be either class A or B).

Case 3—*error-fixing* correction. The analyst found a classification error via image interpretation whereby pixels classified as *pure* class A were judged to be class B. The analyst added new training data for class B labels; for simplicity only one class B training pixel is shown (Fig. 3b). The impact of adding this new training pixel led to multiple splitting of the feature space until the minimum spectral tolerance threshold was met. The partition with the new training pixel was split into sub-partitions (*pure* class A, *pure* class B, and two *unlabeled* partitions). In addition, some surrounding feature space was split into *pure* class A partitions and also some new *unlabeled* partitions were introduced into nearby regions of the feature space where training pixels were sparse.

Case 4—*error-fixing* correction. This case is similar to case 3 but the addition of a class B training pixel resulted in a small *indivisible* partition because the minimum spectral tolerance threshold was met. The pixels falling in this new partition were categorized as *indivisible*.
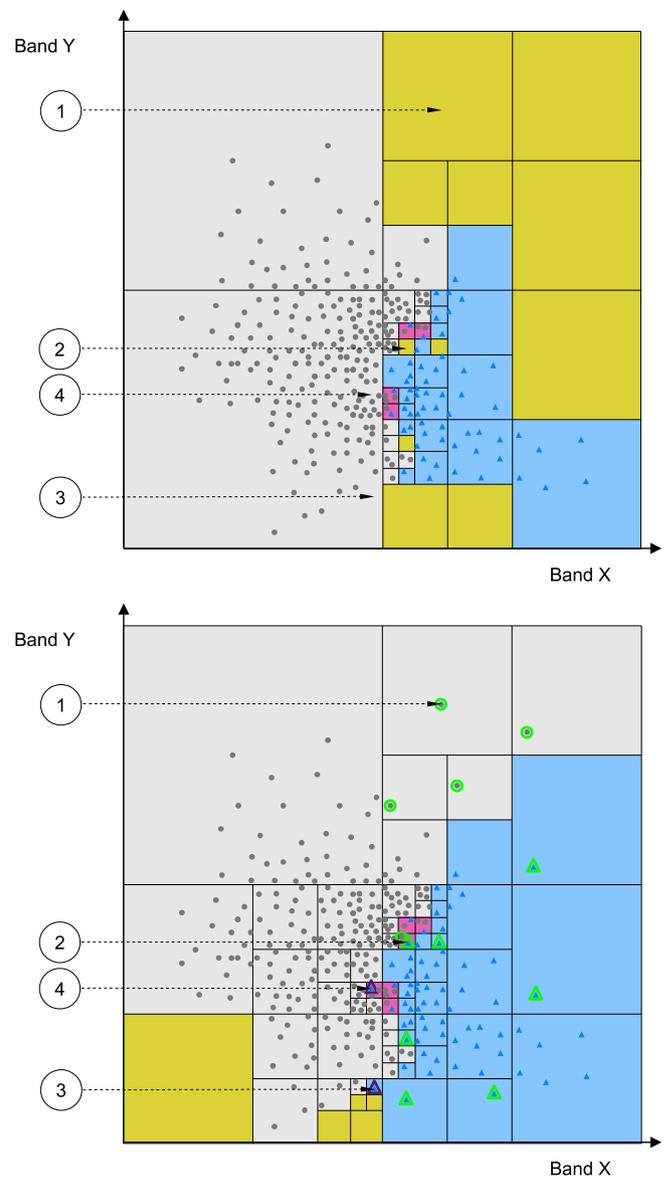

Fig. 3. Partitioning of the feature space before (a) and after (b) new training pixels are added by the analyst (the partitions in a are the same as in Fig. 1f). New training data are shown as outlined dots (green outlines show *gap-filling* and dark magenta ones show *error-fixing* corrections). The black dashed arrows show four specific scenarios (see text for details).

3.4. Final classification post-processing

In this study open surface water (SW) and permanent snow and ice (SI) training pixels across the CONUS were collected and the nested segmentation approach applied. The resulting SW classification had four classes: *SW*, *not SW*, *indivisible* and *unlabeled.* The *SW* and *not SW* classes were derived from *pure* feature space partitions. Similarly, the SI classification had four classes: *SI*, *not SI*, *indivisible*, and *unlabeled.* In addition, class probabilities were stored that are similar to the per pixel class probabilities provide by CART classifiers. Conventional CART algorithms recursively partition training data into more homogeneous subsets referred to as nodes (Breiman, 2001). The probability of class membership for each node is defined as the portion of training pixels of the class in the node, and the probability of class membership for each classified pixel is assigned by the node probability (Breiman, 2001). Each CART node is equivalent to a nested segmentation feature space partition (e.g., a hypercube shown in Fig. 3b). It is reasonable therefore to compute the probability of class membership in the same way. Thus, the probability of SW was computed for each partition as the number of training pixels of class *SW* divided by the total number of training pixels in the partition. The probability of SI was computed for each partition in the same manner. For convenience, the probability values were multiplied by 100 to give percentages rounded to the nearest integer. In this way, the pixels classified as *SW* and *SI* had 100% probability, the pixels classified as *not SW* and *not SI* had 0% probability, the pixels classified as *indivisible* had class probabilities in the range 1% to 99%.

Some post-classification heuristics were applied to reduce commission errors. The 0.0002777° National Elevation Data (NED) (Gesch, 2007) first derivative slope product was reprojected to 30 m and all pixels where slopes were >4° were reclassified as *not SW*. This was based on the assumption that water would not be present on slopes (Bwangoy, Hansen, Roy, De Grandi, & Justice, 2010). The median 5-year high gain brightness temperature (10.40–12.50 μm) was used to identify locations likely to be too warm for persistent snow and ice accumulation. An empirical examination found that locations with a median 5-year high gain brightness temperature of 20 °C provided a conservative threshold and all pixels in the SI classification with brightness temperature above this threshold were reclassified as *not SI.*

4. Results

4.1. Training data selection

To create an initial training data set, 124 and 54 unambiguous training pixels for SW and SI characterizations, respectively, were collected across the CONUS by examination of the WELD weekly data. Care was taken to select only pure class training pixels. Subsequently, in the iterative nested segmentation approach for the SW classification, pixels containing no water were considered as training class *not SW* (this corresponds to class A on pictures 1, 2 and 3) and pixels containing any portion of water (>0%) were taken as class *SW* (i.e., class B on pictures 1, 2 and 3). Similarly, for the SI classification, pixels containing no snow or ice were considered as training class *not SI* and pixels fully or partially covered by snow (>0%) were taken as class *SI*. To ensure representative class variation, training pixels were purposefully collected across the CONUS. Only cloud and shadow free training data were selected. The training data class labels were checked visually using the "Open in Google Earth" tool (http://gis-lab.info/qa/open-in-google-en.html) which allowed a comparison with high spatial resolution near-contemporaneous GoogleEarth™ airborne imagery. After several iterations of the supervised active learning nested segmentation process, a total of 296,363 and 93,496 training pixels for the SW and SI characterizations, respectively, were collected.

4.2. CONUS classification

Browse images of the final open surface water (SW) and permanent snow and ice (SI) classification results are shown in Figs. 4 and 5 respectively. The classification results are shown superimposed on a false color image of the 5-year median metrics Landsat bands 5, 4, and 3, to provide geographic context. The transparent areas correspond to the pixel locations classified as *not SW* and *not SI*. For both data sets a total of 9,976,500,374 30 m pixels were classified. As it is not possible to visualize all of the CONUS at 30 m resolution in a single image (Roy et al., 2010) the browse classification images were generated by labelling a reduced resolution browse image pixel as *SW* or *SI* if any of the underlying 30 m pixels were classified as these classes. This necessarily overemphasizes the spatial distribution of the *SW* and *SI* classes.
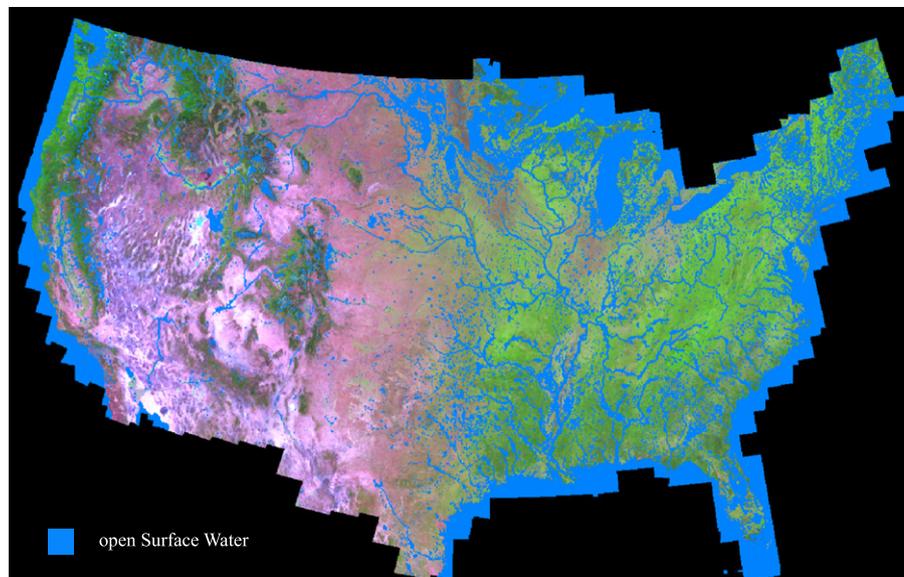

Fig. 4. Open surface water classification superimposed over 5-year median metrics (Landsat bands 5, 4, 3 shown as red, green, blue), Albers equal area projection. *Indivisible* and *unlabeled* pixels are shown as *SW* (blue color).
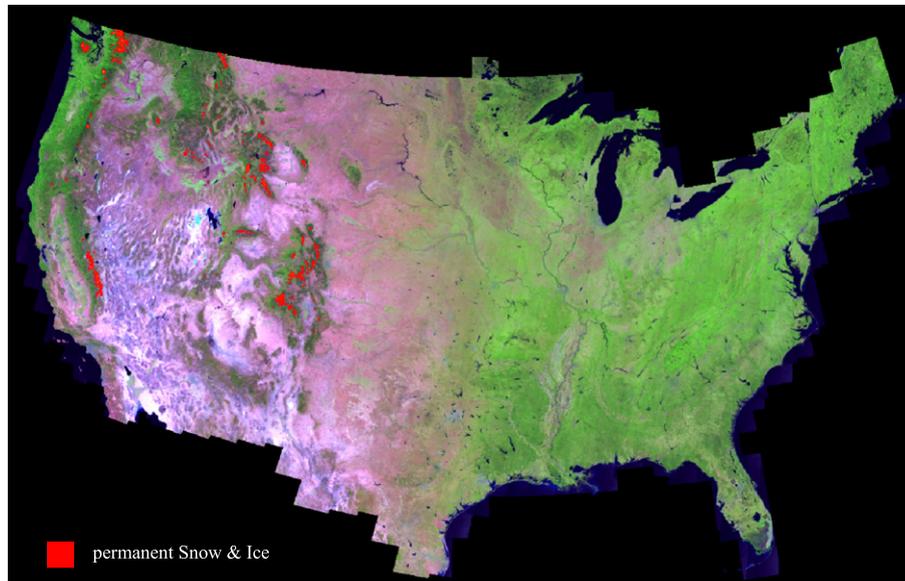
Fig. 5. Permanent snow and ice classification superimposed over 5-year median metrics (Landsat bands 5, 4, 3 shown as red, green, blue), Albers equal area projection. *Indivisible* and *unlabeled* pixels are shown as *SI* (red color).

SW commission errors were often found in mountainous areas. For example, deep shadows on north-facing slopes have very low reflectance and are often classified as water. However, applying the post-classification slope heuristic removed the majority of these errors. For the SI classification, the post-classification temperature heuristic removed highly reflective salt pans that are spectrally similar to SI in the reflective wavelengths. Known omission errors in SW classification are related to the date of NED data set derivation, which varies across CONUS from 1923 to 2013 (Stoker, Heidemann, Evans, & Greenlee, 2013). For the SI classification, omission errors occur along the edges of some of the snow covered areas, likely because the thermal band data are sensed at 60 m and not at the 30 m resolution defining the reflective wavelength bands.

Table 1 summarizes the percentage of the CONUS 30 m pixels classified into the different classes for the two classifications. The percentage of pixels belonging to the *indivisible* category is insignificant (0.129% and 0.007% for the SW and SI classifications respectively). The *indivisible* pixels in the SW classification included pixels with shadows occurring more than 50% of the time in the weekly WELD data, typically on urban and impervious surfaces. Other *indivisible* land cover types confounding water discrimination included volcanic rocks and exposed soil surfaces such as Belknap crater, OR, and Sunset crater, AZ. The majority of *indivisible* pixels in the SI classification were located on salt pans with high visible and infrared reflectance. The percentage of pixels belonging to the *unlabeled* category was even smaller (0.042% and 0.001% for the SW and SI classifications) and is indicative of the efficacy of the nested segmentation classification approach.

Table 1
Percentage of the number of CONUS pixels (out of a total of 9,976,500,374 30 m pixels considered) that were classified into the different classes (SW = open surface water, not SW = not open surface water, SI = permanent snow and ice, not SI = not permanent snow and ice).

CONUS open surface water classification percentages			
SW	Not SW	Indivisible	Unlabeled
9.807	90.021	0.129	0.042

CONUS permanent snow and ice classification percentages			
SI	Not SI	Indivisible	Unlabeled
0.060	99.932	0.007	0.001

A total of 9.8% of the CONUS pixels were classified as *SW* (Table 1). The spatial distribution of the *SW* class (Fig. 4) appear generally coherent with the major lakes, rivers, inland water bodies, and near shore oceans. Only 0.06% of the CONUS pixels were classified as *SI* (Table 1). The *SI* class (Fig. 5) occurs only in high altitude snow prone areas and principally depicts the extent of glaciers within CONUS (Barnes & Roy, 2010; Krimmel, Key, Fagre, & Menicke, 2002). The SI classification was derived using 5-year median metrics defined from the growing season, specifically week 16 to week 46 or the median spectral signature from April 15th to November 17th. This signature represents close to minimum snow and ice coverage for the growing season, mainly that of glaciers.

4.3. Open surface water classification comparison

To estimate the quality of the WELD SW classification, a pixel by pixel comparison with two recent, similar national scale products was undertaken: the vector Shuttle Radar Topography Mission (SRTM) water body data set (SWBD, 2005) and the 30 m National Land Cover Database (NLCD 2006) Open Water class (Fry et al., 2011). To our knowledge, the SRTM water body data have not been formally validated, while the NLCD water class has a 93% map accuracy (Wickham et al., 2013). The SWBD was reprojected to the WELD Albers projection and rasterized to a 30 m pixel size in the WELD pixel grid. Class 11 (Open Water) of the NLCD 2006 land cover product was considered as open surface water (*SW*) and the other classes were considered as not open surface water (*not SW*).

The WELD CONUS open surface water map was compared with the SWBD and NLCD data to generate two-way confusion matrices. Conventional accuracy statistics (Cohen's Kappa, user's, producer's and overall accuracies) were then derived from the confusion matrices (Foody, 2002). In this analysis pixels belonging to the *indivisible* and *unlabeled*

Table 2
Confusion matrix comparison of WELD open surface water (SW) 30 m classification (Fig. 4) with SRTM Water Body data (SWBD). *SW* = open surface water, *not SW* = not open surface water. Overall accuracy: 99.4; Cohen's Kappa: 0.966.

	WELD not SW	WELD SW	User's accuracy
SWBD not SW	8959628989	38378857	99.6
SWBD SW	21357582	957134946	97.8
Producer's accuracy	99.8	96.1	

Table 3

Confusion matrix comparison of WELD open surface water (SW) 30 m classification (Fig. 4) with NLCD open water map. SW = open surface water, not SW = not open surface water. Overall accuracy: 99.4; Cohen's Kappa: 0.935.

	WELD not SW	WELD SW	User's accuracy
NLCD not SW	8489643813	23405746	99.7
NLCD SW	32922440	427686146	92.9
Producer's accuracy	99.6	94.8	

categories were considered to be *SW* as they usually occur at the edge of water bodies (this is illustrated for example in Fig. 2 left) and their inclusion as *SW* is deemed negligible as < 0.17% of the CONUS were *indivisible* or *unlabeled* (Table 1).

Tables 2 and 3 illustrate the high level of agreement between the SW classification and the SWBD and NLCD data respectively. All

9,976,500,374 classified WELD pixels were compared with the SWBD product; 8,973,658,145 classified WELD pixels were compared with the NLCD 2006 data as NLCD covers a smaller area than the WELD data. The Cohen's Kappa (Cohen, 1960) coefficient was 0.97 and 0.94 for SW vs. SWBD and SW vs. NLCD respectively; the overall percent correct classification accuracy in both of cases was greater than 99%. The user's accuracies for the *SW* class were 97.82% (SW vs. SWBD) and 92.86% (SW vs. NLCD) and the producer's accuracies were 96.14% and 94.81% respectively, reflecting the high level of agreement between the maps.

4.4. Nested feature space partitioning analysis

Fig. 6 illustrates the final open surface water (SW) feature space partitioning, i.e. the SW classifier. A total of four metrics, the median
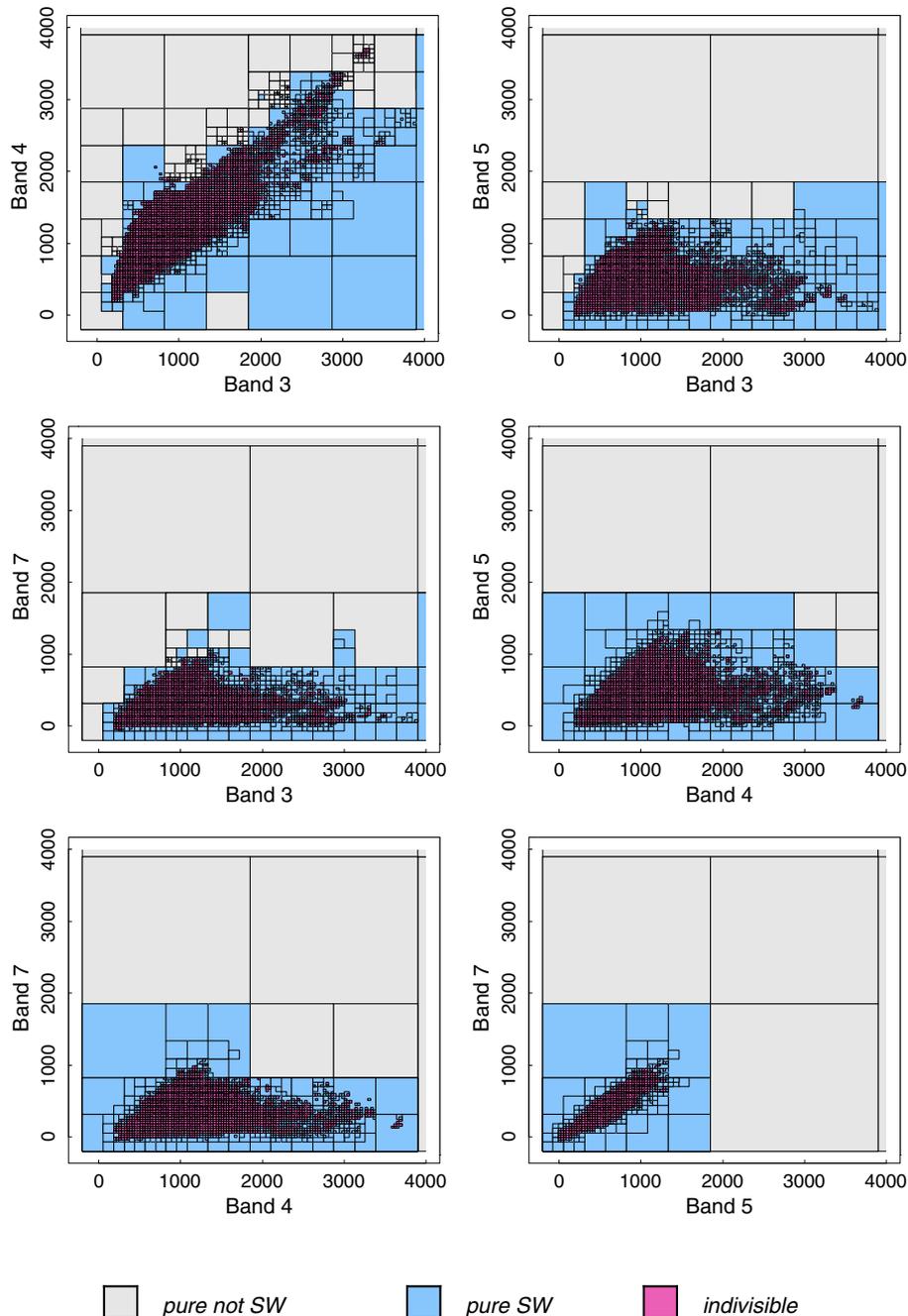

Fig. 6. Spectral scatterplots of the final partitioning of the feature space for the open surface water (SW) classification derived from a final total of 296,363 training pixels. Only the *pure SW* (blue), *pure not SW* (gray), and *indivisible* partitions (magenta) are illustrated.

Table 4
Final SW model and distribution by category (pure, indivisible and unlabeled) of feature space partitions, feature space volume and CONUS classified pixels, expressed in absolute numbers (#) and percentage (%).

Category	Partitions, #	Partitions, %	Volume, %	Pixels, #	Pixels, %
Pure	67538	45.003	10.59751	9959408985	99.829
Indivisible	16262	10.836	0.00003	12896672	0.129
Unlabeled	66276	44.162	89.40246	4194717	0.042

5 year reflectance derived for bands 3 (0.63-0.69 μm), 4 (0.78-0.90 μm), 5 (1.55-1.75 μm), and 7 (2.09-2.35 μm), were classified. As it is not possible to visualize a four dimensional feature space, the six possible combinations of two dimensional feature spaces are shown. Due to visualization complexity only the *indivisible* (magenta) and *pure* (SW in blue, *not SW* in gray) partitions are illustrated. This illustration over-emphasizes the extent of *indivisible* partitions, although in the four-dimensional feature space they covered only a small fraction of the feature space volume (Table 4).

Tables 4 and 5 summarize the number (#) and percentage (%) of the feature space partitions of different categories (*pure*, *indivisible*, or *unlabeled*) and also the percentage volume of feature space they occupy for the open surface water (SW) and permanent snow and ice (SI) classifications respectively. In addition, the last two columns of these tables summarize the number and percentage of classified pixels of the three different categories. The number of *unlabeled* partitions and associated feature space volume occupied is considerable for both classifications, but these partitions occupy a "sparse" or "empty" volume of the feature space, where only isolated pixels are located. More than 99% of pixels in the SW and SI classification were *pure* and only 0.042% and 0.001% of pixels in the SW and SI classifications respectively were *unlabeled*. The *indivisible* partitions covered only a very small portion of the feature space: 0.00035% (SW) and 0.002% (SI) with only a minor percentage of pixels categorized as *indivisible* in classified images: 0.129% (SW) and 0.007% (SI).

Tables 6 and 7 summarize by partition size (partition side length in units of reflectance × 10,000) the number, percentage and volume occupied by the partitions for the final SW and SI classifications respectively. The respective 32 and 128 (which correspond to 0.0032 and 0.0128 defined in reflectance units with a 10,000 scaling factor) minimum spectral tolerance thresholds employed resulted in nine splits for the SW and seven splits for the SI models, respectively. The hyper-

Table 5
Final SI model and distribution by category (pure, indivisible and unlabeled) of feature space partitions, feature space volume and CONUS classified pixels, expressed in absolute numbers (#) and percentage (%).

Category	Partitions, #	Partitions, %	Volume, %	Pixels, #	Pixels, %
Pure	13550	47.244	24.185	9975718866	99.992
Indivisible	4721	16.460	0.002	696067	0.007
Unlabeled	10410	36.296	75.813	85441	0.001

Table 6
Final open surface water (SW) classification feature space number (#) and percentage (%) of partitions of different sizes and the percentage volume of feature space occupied.

Edge	Partitions, #	Partitions, %	Volume, %
8192	15	0.010	93.7500
4096	14	0.009	5.4688
2048	27	0.018	0.6592
1024	57	0.038	0.0870
512	263	0.175	0.0251
256	1219	0.812	0.0073
128	5427	3.616	0.0020
64	23726	15.809	0.0006
32	119328	79.512	0.0002

Table 7
Final permanent snow and ice (SI) classification feature space number (#) and percentage (%) of partitions of different sizes and the percentage volume of feature space occupied.

Edge	Partitions, #	Partitions, %	Volume, %
8192	14	0.049	87.500
4096	25	0.087	9.766
2048	89	0.310	2.173
1024	286	0.997	0.436
512	946	3.298	0.090
256	4425	15.428	0.026
128	22896	79.830	0.009

volumes, occupied by partitions of different sizes, are very unequal; for example, only 29 partitions had side lengths of 8192 and 4096 scaled reflectance in the SW classification (Table 6). However, these partitions cover 99.2% of the feature space volume. Conversely, less than 0.0002% of the feature space volume is occupied by partitions with side lengths of 32 (the minimum spectral tolerance threshold). Similar results were found for the SI classification (Table 7).

4.5. Nested segmentation and CART classifiers comparison

To provide confidence in the nested segmentation algorithm the training data used to generate the final SW classification were used again to generate a SW classification but with a standard bagged CART classifier (Breiman et al., 1984; Hansen, 2012). Twenty-five bagged classification trees were generated; each time 10% of the training data were sampled at random with replacement and used to generate a tree. Tree growth was terminated when additional splits decreased model deviance by less than 0.001 of the root node deviance. Each pixel was classified 25 times using 25 bagged classification trees. All per pixel results were ranked over the 25 trees and the median water class membership probability was taken as the final result. Pixels with probability ≥50% and <50% were considered to be the open surface water (WS) and the not open surface water (non WS) classes, respectively. The final classification post-processing (Section 3.4) was applied to the CART classification to make it comparable with the nested segmentation classification.

Table 8 shows the CONUS confusion matrix summarizing the two SW classifications, assuming that the CART SW classification is "truth". These results indicate a high overall classification correspondence (Cohen's Kappa coefficient 0.99, overall percent correct 99.9%) with user's and producer's accuracies of 92.36% and 94.59% respectively.

Fig. 7 shows detailed examples comparing the SW classifications provided by the nested segmentation (left column), the CART classification (middle column) and the median of Landsat bands 5, 4 and 3 (right column) shown to provide geographic context. The top two rows illustrate examples where the two classification results are in evident agreement for extensive open water bodies (Louisiana) and more spatially complex prairie pot hole lakes (South Dakota). The bottom row shows an example where the two SW classifications disagree markedly. Examination of this classification difference (by inspection of Google Earth high spatial resolution data) indicates that it is due to CART commission errors occurring over an extensive area of mining deposits close

Table 8
Confusion matrix comparison of two open surface water (SW) 30 m classifications using the same training data sets, the same set of metrics and two different classifiers—25 bagged trees (CART) and nested segmentation (NS). SW = open surface water, not SW = not open surface water. Overall accuracy: 99.9; Cohen's Kappa: 0.99.

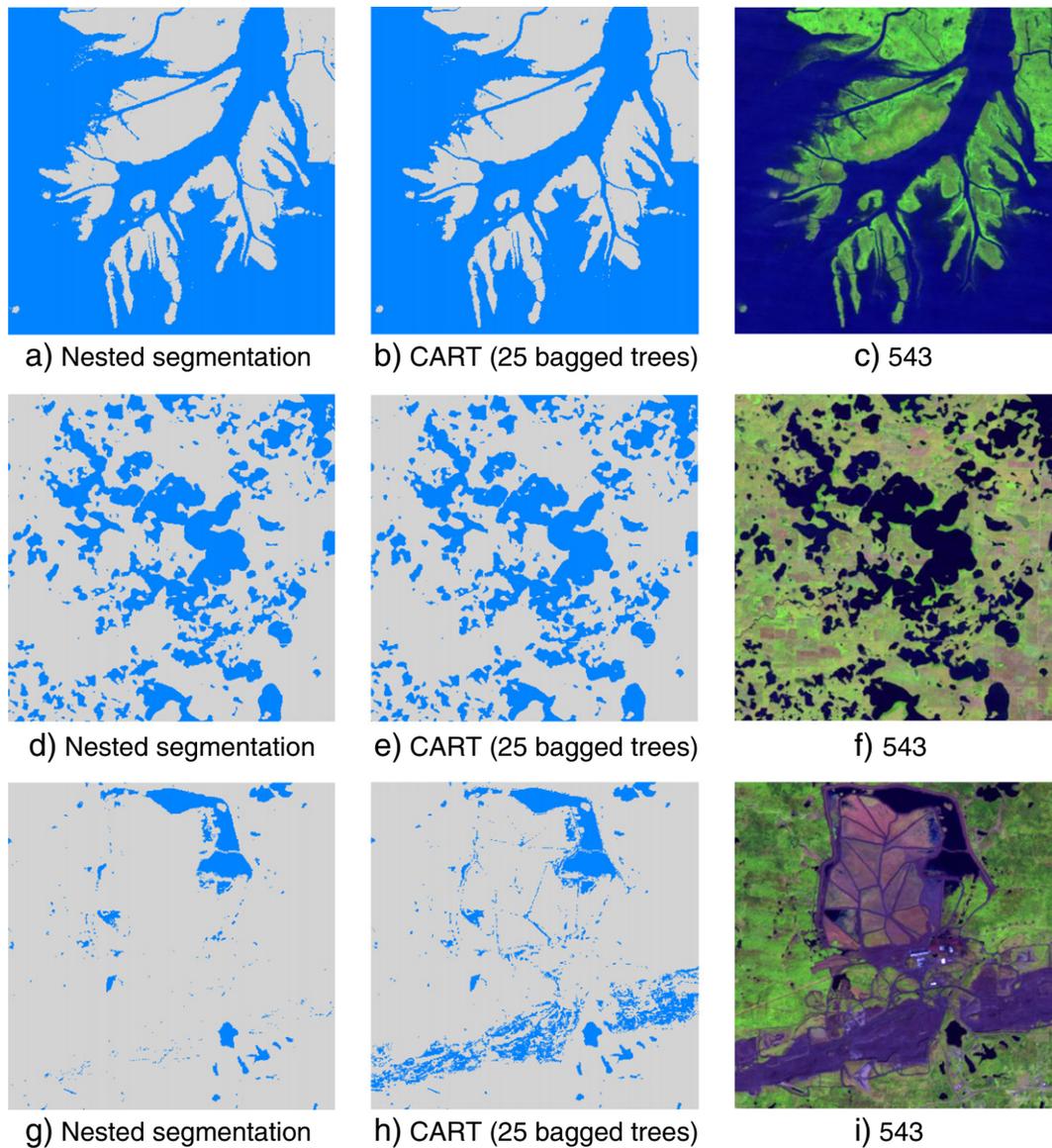|	NS not SW	NS SW	User's accuracy
CART not SW	8974620567	4084024	99.95
CART SW	6366004	991429779	92.36
Producer's accuracy	99.93	94.59	

Fig. 7. A detailed 400 × 400 30 m pixel comparison of nested segmentation classification results (left column) and bagged CART (25 tree) classification (middle column) derived using the same Landsat metrics and the same training data derived by application of the nested segmentation guidance procedure. The right column shows the 5-year median of bands 5, 4, 3 as RGB for geographic reference. Top row: Louisiana, 29°26'49.24"N, 91°18'8.01"W, Middle row: South Dakota, 45°38'36.10"N, 97°30'25.01"W, Bottom row: Minnesota, 47°34'14.00"N, 92°38'39.15"W.

to Mountain Iron, Minnesota. These results provide confidence that the image-interpreted nested segmentation approach is quite robust.
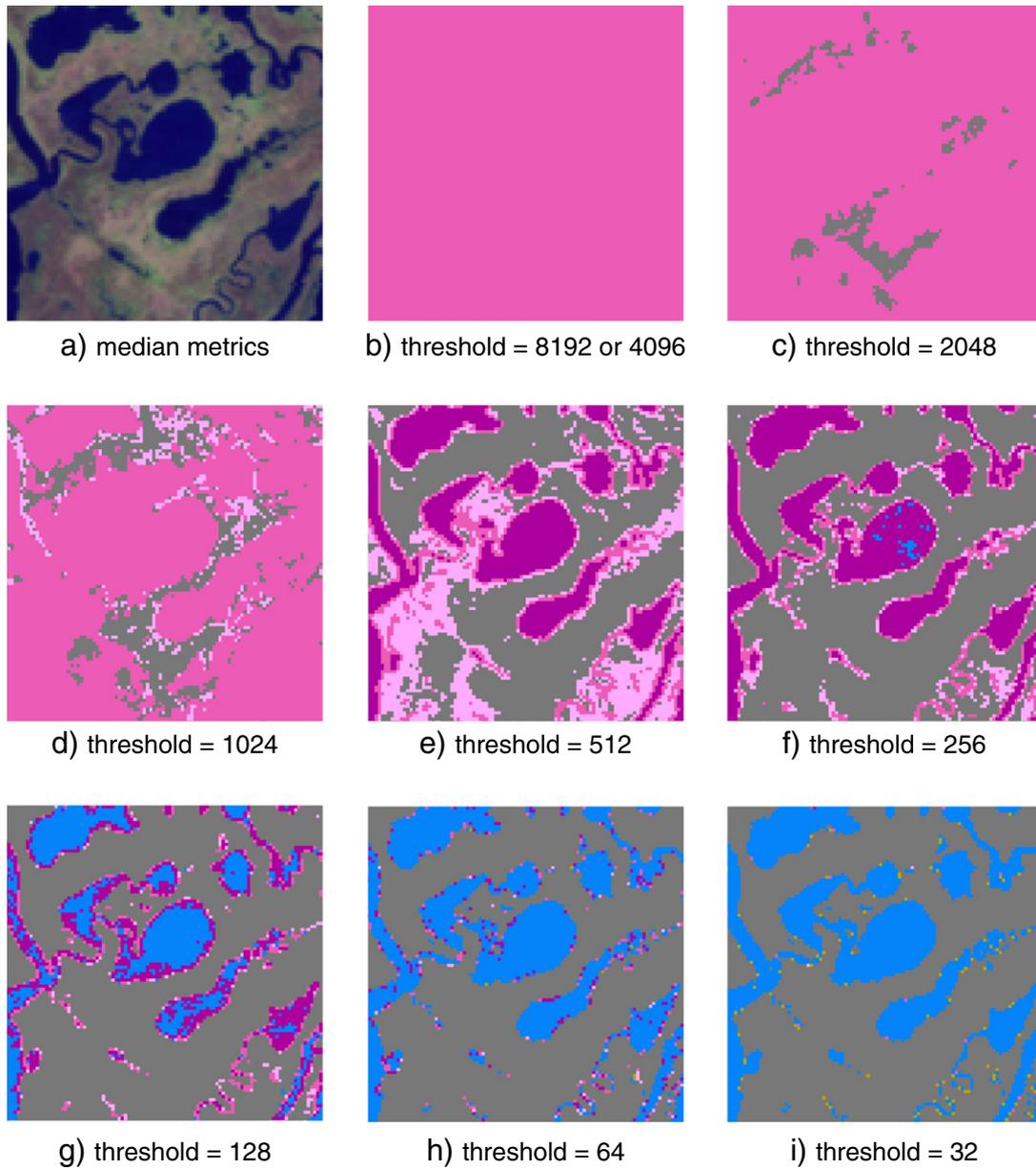
5. Discussion

5.1. Analyst interpretation workload

Achieving a reliable classification is a function of the analyst's interpretation skills and the ability to recognize when diminishing returns indicate a fundamental limitation in improving the map characterization. The only preset parameter is the minimum spectral tolerance threshold value, which if set too low can result in interminable interpretation of partitions. Each *pure* partition and each *indivisible* partition is defined by at least one and two training pixels respectively. The more partitions, the more detailed training and the more labor-intensive the task becomes. Conversely, increasing the minimum spectral tolerance threshold leads to an exponential reduction of manual work, but also to a possible loss of quality in the map output by retaining a relatively larger transition zone in the final product. A balance is sought between

the amounts of labor performed in interpreting/iterating the product versus final map quality/accuracy.

Fig. 8 demonstrates how varying the minimum spectral tolerance threshold affects the classification quality. With a large threshold (8192 and 4096, Fig. 7b), no classification is possible—all classified pixels are categorized as *indivisible*. Using a threshold of 256 (Fig. 7f) enables only core areas of water and land to be identified. Employing a minimum spectral tolerance threshold of 32 leads to almost a clear open surface water characterization with an insignificant number of *unlabeled* and *indivisible* pixels in the final result (Fig. 7i).

The nested segmentation algorithm reduces overall effort during training data set creation. As result, the training data volume used in nested segmentation is relatively small. For example, in our previous CONUS research using the CART classifier, training data set consisted of 112,489,590 pixels for open surface water classification, 10,912,417 pixels for percent Tree Cover, 151,025,252 pixels for percent of Bare Ground (Hansen et al., 2011), 1,515,582 pixels for Forest Cover Loss and 12,589,299 pixels for Bare Ground Gain (Hansen et al., 2014). The size of the training set used to build the nested segmentation

Fig. 8. Example product sequence illustrating the effect of minimum spectral tolerance threshold from a SW example. The same training set and the same set of image metrics were used in all examples. Only the minimum spectral tolerance threshold was changed. For the image reference in (a), 5-year median metrics are shown (Landsat bands 5, 4, 3 as red, green, blue).

model for SW consisted of only 296,363 training pixels; for SI, 93,496 training pixels. Nested segmentation results in a targeted and comparatively small training data requirement compared to traditional approaches.

5.2. Known issues and limitations

The nested segmentation algorithm has some limitations. First, due to the equilateral shape of the partitions, the input data should be normalized, i.e. have similar dynamic ranges across input variables. WELD

weekly mosaics are normalized to top of atmosphere reflectance with a valid range from 0 to 10,000 for all bands with some allowance due to known uncertainties (Markham & Helder, 2012). Converting WELD weekly data into median metrics does not change the range of valid values. However, adding an ancillary layer of another physical variable, for example NED-derived slope, varying from 0° to 90°, would not conform to the spectral splitting rule used in this classification.

Another disadvantage of the approach is the challenge of processing a large number of metrics. Each split divides k mixed parent partitions into $k*2^n$ child partitions (where n is a number of metrics). Increasing

the dimensionality of the feature space leads to an exponential increase in the amount of training data required to cover all partitions, an effect commonly referred to as the curse of dimensionality (Bellman, 1954; Hughes, 1968). In this research only four metrics were sufficient to achieve nearly complete separation of the two classes. For more complicated thematic targets (e.g., forests or croplands), more metrics would likely be needed to achieve a viable classification accuracy. A modification of the partitioning process would be required in order to use this method with a larger number of metrics.

Finally the method has a high sensitivity to errors in the training data set. A mislabeled training pixel may launch multiple unnecessary splits of a *pure* partition, which should not be split. This leads to the further appearance of *unlabeled* partitions and multiple *unlabeled* pixels in a classified image. However, using a relatively small training data set, mislabeled pixels are easily discovered as they will in effect behave as outliers and result in spurious partitions of the feature space.

5.3. Advantages and future modifications of the method

A random sampling approach at the stage of training data set collection does not ensure the creation of a viable training data set, as it does not target the spectral frontier along class boundaries. The nested segmentation partitioning algorithm identifies areas which already contain training data (*pure* and *indivisible* pixels and partitions), avoiding needless training duplication, extra work and computation. By facilitating the direct identification and minimization of the transition zone between two classes, the nested segmentation algorithm maximizes the spectral volume occupied by pure partitions. Conversely, *unlabeled* pixels and partitions serve as a query from classifier to analyst, explicitly highlighting untrained spectral volumes. These regions are coded and mapped, enabling their subsequent investigation and interpretation (gap-filling correction as in Fig. 3). Most modern classifiers adapted to remote sensing problems do not provide any information about areas that lack training data. The parallelepiped classifier (Richards, 1999) is one that does provide information on spectral regions not labeled by training data.

Future modification of the classifier will focus in three directions: 1) a modification of the partitioning process to allow for the use of more features, 2) testing models with other land cover themes, in other regions and in other time frames, and 3) an automatic calculation of minimum spectral tolerance for different types of land cover. Progress on these three aspects will enable testing the advanced nested segmentation with land cover types requiring a richer feature space, for example forest cover. Future model implementations will be tested with WELD data across all Landsat 5/7 epochs (since 1985), providing a means to document change over time using the nested segmentation approach.

Though the nested segmentation algorithm has been developed only for binary classification, there are no technical limits to building an algorithm for the partitioning of a feature space for three or more predefined classes. Multiclass classification can be implemented for both a) combined use of multiple binary classifiers and b) single model for multiple classes, where a probability of each class can be assigned to each partition. The potential challenge concerns the curse of dimensionality (Bellman, 1954; Hughes, 1968) for complicated thematic classes (e.g., vegetation), where the exponential growth of training samples can be expected. This topic will be the focus of forthcoming research on the nested segmentation approach.

6. Conclusions

We developed and applied a novel active learning classifier, which we call *nested segmentation*, to CONUS multi-temporal Landsat data. Active learning as implemented in our approach enables guided iterations of the map product through 1) identifying spectral regions that lack training data, 2) identifying transition zones between the two classes

of interest, and 3) reducing the transition zone to maximize the identification of spectral regions consisting of a single land cover class.

The result is a training data set that is 1) *representative* of relevant intraclass spectral variation. In other words, the full extent of each class's spectral signature in hyper-dimensional space is targeted for training. The training data set is also 2) *concentrated* in the regions of spectral confusion between the classes, or the transition zones. By placing more effort in refining the spectral boundaries through concentrated training data derivation, the region of *indivisible* pixels is reduced by maximizing the delineation of pure spectral space belonging to a single class. Nested segmentation is best suited to situations where labels are scarce and very difficult, time-consuming, or expensive to obtain. Given a competent image analyst, high fidelity land cover maps should be easily realized using nested segmentation.

The products described here are part of the WELD land cover data sets for the CONUS for the 2006–2010 Landsat 7 epoch and are available for free download from a United States Geological Survey (USGS) server (http://e4ftl01.cr.usgs.gov/WELD/LCLUC/).

Acknowledgments

References

Arora, M.K., & Foody, G.M. (1997). Log-linear modeling for the evaluation of the variables affecting the accuracy of probabilistic, fuzzy and neural network classifications. *International Journal of Remote Sensing, 18*, 785–798.

Barnes, C.A., & Roy, D.P. (2010). Radiative forcing over the conterminous United States due to contemporary land cover land use change and sensitivity to snow and interannual albedo variability. *Journal of Geophysical Research, 115*, G04033. http://dx.doi.org/10.1029/2010JG001428.

Bellman, R. (1954). The theory of dynamic programming. *Bulletin of the American Mathematical Society, 60*, 503–516.

Bernard, A.C., Wilkinson, G.G., & Kanellopoulos, I. (1997). Training strategies for neural network soft classification of remotely sensed imagery. *International Journal of Remote Sensing, 18*, 1851–1856.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees.* Monterey, CA: Wadsworth and Brooks/Cole.

Broich, M., Hansen, M.C., Potapov, P., Adusei, B., Lindquist, E., & Stehman, S.V. (2011). Time-series analysis of multi-resolution optical imagery for quantifying forest cover loss in Sumatra and Kalimantan, Indonesia. *International Journal of Applied Earth Observation and Geoinformation, 13*, 277–291.

Bwangoy, J.R., Hansen, M.C., Roy, D.P., De Grandi, G., & Justice, C.O. (2010). Wetlands mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices. *Remote Sensing of Environment, 114*, 73–86.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Cortes, C., & Vapnik, V.N. (1995). Support-vector networks. *Machine Learning, 20*.

Dagan, I., & Engelson, S. (1995). Committee-based sampling for training probabilistic classifiers. *Proceedings of the International Conference on Machine Learning (ICML)* (pp. 150–157). Morgan Kaufmann.

DeFries, R.S., Field, C.B., Fung, I., Justice, C.O., Los, S., Matson, P.A., et al. (1995). Mapping the land surface for global atmosphere-biosphere models: Towards continuous distributions of vegetation's functional properties. *Journal of Geophysical Research, 100*, 20867–20882.

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*(2), 179–188.

Fix, E., & Hodges, J.L., Jr. (1951). Discriminatory analysis, nonparametric discrimination: Consistency properties. *USAF School of Aviation Medicine, Randolph Field, Tex., Project 21-49-004, Rept. 4, Contract AF41(128)-31, Texas, February 1951.*

Foody, G.M. (1999). The significance of border training patterns in classification by a feedforward neural network using back propagation learning. *International Journal of Remote Sensing, 20*, 3549–3562.

Foody, G. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80*, 185–201.

Foody, G.M., & Mathur, A. (2004a). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing, 42*, 1335–1343.

Foody, G.M., & Mathur, A. (2004b). Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification. *Remote Sensing of Environment, 93*, 107–117.

Foody, G.M., & Mathur, A. (2006). The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment, 103*, 179–189.

Foody, G.M., McCulloch, M.B., & Yates, W.B. (1995). The effect of training set size and composition on artificial neural network classification. *International Journal of Remote Sensing, 16*, 1707–1723.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., et al. (2011). Completion of the 2006 National Land Cover Database for the Conterminous United States. *PE&RS, 77*(9), 858–864.

Fujii, A., Tokunaga, T., Inui, K., & Tanaka, H. (1998). Selective sampling for example based word sense disambiguation. *Computational Linguistics, 24*(4), 573–597.

Gesch, D.B. (2007). The national elevation dataset. In D. Maune (Ed.), *Digital elevation model technologies and applications: The DEM users manual* (pp. 99–118) (2nd ed.). Bethesda, Maryland: American Society for Photogrammetry and Remote Sensing.

Goward, S.N., Masek, J.G., Williams, D.L., Irons, J.R., & Thompson, R.J. (2001). The Landsat 7 mission, Terrestrial research and applications for the 21st century. *Remote Sensing of Environment, 78*, 3–12.

Hansen, M.C. (2012). Classification trees and mixed pixel training data. In C. Giri (Ed.), *Remote sensing of land cover: Principles and applications.* New York: Taylor and Francis.

Hansen, M.C., Egorov, A., Potapov, P.V., Stehman, S.V., Tyukavina, A., Turubanova, S.A., et al. (2014). Monitoring conterminous United States (CONUS) land cover change with Web-Enabled Landsat Data (WELD). *Remote Sensing of Environment, 140*, 466–484.

Hansen, M.C., Egorov, A., Roy, D.P., Potapov, P., Ju, J., Turubanova, S., et al. (2011). Continuous fields of land cover for the conterminous United States using Landsat data: First results from the Web-Enabled Landsat Data (WELD) project. *Remote Sensing Letters, 2*, 279–288.

Hansen, M.C., & Loveland, T.R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment, 2012*, 66–74.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., et al. (2013). High-resolution global maps of 21st century forest cover change. *Science, 342*(6160), 850–853.

Hansen, M.C., Stehman, S.V., Potapov, P.V., Loveland, T.R., Townshend, J.R.G., DeFries, R.S., et al. (2008). Humid tropical forest clearing from 2000 to 2005 quantified using multi-temporal and multi-resolution remotely sensed data. *Proceedings of the National Academy of Sciences, 105*, 9439–9444.

Hauptmann, A., Lin, W., Yan, R., Yang, J., & Chen, M.Y. (2006). Extreme video retrieval: Joint maximization of human and computer performance. *Proceedings of the ACM Workshop on Multimedia Image Retrieval* (pp. 385–394). ACM Press.

Hoi, S.C.H., Jin, R., & Lyu, M.R. (2006). Large-scale text categorization by batch mode active learning. *Proceedings of the International Conference on the World Wide Web* (pp. 633–642). ACM Press.

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory, 14*, 55–63.

INPE (2013). http://www.obt.inpe.br/prodes/index.php (Last accessed October 1st 2013)

Jackson, Q., & Landgrebe, D. (2001). An adaptive classifier design for high-dimensional data analysis with a limited training data set. *IEEE Transactions on Geoscience and Remote Sensing, 39*(12), 2664–2679.

Johnson, D.M., & Mueller, R. (2010). The 2009 cropland data layer. *Photogrammetric Engineering and Remote Sensing, 76*, 1202–1205.

Ju, J., & Roy, D.P. (2008). The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sensing of Environment, 112*, 1196–1211.

Jun, G., & Ghosh, J. (2008). An efficient active learning algorithm with knowledge transfer for hyperspectral data analysis. *IEEE Geosci. Remote Sens. Symp. IGARSS.*

Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics, 43*(1), 59–69.

Kohonen, T., & Honkela, T. (2007). *Kohonen network.* Scholarpedia.

Krimmel, R.M., Key, C.H., Fagre, D.B., & Menicke, R.K. (2002). Glaciers of the conterminous United States—Glaciers of the western United States. In R.S. WilliamsJr., & J.G. Ferrigno (Eds.), *Satellite image Atlas of glaciers of the world: North America. United States Geological Survey Professional paper 1386-J-1.* US Government Printing.

Krishnamurthy, V. (2002). Algorithms for optimal scheduling and management of hidden Markov model sensors. *IEEE Transactions on Signal Processing, 50*(6), 1382–1397.

Lang, K., & Baum, E. (1992). Query learning can work poorly when a human oracle is used. *Proceedings of the IEEE International Joint Conference on Neural Networks* (pp. 335–340). IEEE Press.

Lewis, D., & Catlett, J. (1994). Heterogeneous uncertainty sampling for supervised learning. *Proceedings of the International Conference on Machine Learning (ICML)* (pp. 148–156). Morgan Kaufmann.

Li, J., Bioucas-Dias, J.M., & Plaza, A. (2010). Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing, 48*(11), 4085–4098.

Licciardi, G., Pacifici, F., Tuia, D., Prasad, S., West, T., Giacco, F., et al. (2009). Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest. *IEEE Transactions on Geoscience and Remote Sensing, 47*(11), 3857–3865.

Lippitt, C.D., Rogan, J., Li, Z., Eastman, R.J., & Jones, T.G. (2008). Mapping selective logging in mixed deciduous forest: A comparison of machine learning algorithms. *Photogrammetric Engineering and Remote Sensing, 74*, 1201–1212.

Liu, Y. (2004). Active learning with support vector machine applied to gene expression data for cancer classification. *Journal of Chemical Information and Computer Sciences, 44*, 1936–1941.

Markham, B.L., & Helder, D.L. (2012). Forty-year calibrated record of earth-reflected radiance from Landsat: A review. *Remote Sensing of Environment, 122*, 30–40.

Mather, P.M. (2004). *Computer processing of remotely-sensed images. An introduction* (3rd ed.). Chichester, United Kingdom: Wiley (324 pp.).

McCallum, A., & Nigam, K. (1998). Employing EM in pool-based active learning for text classification. *Proceedings of the International Conference on Machine Learning (ICML)* (pp. 359–367). Morgan Kaufmann.

Pal, M., & Mather, P.M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment, 86*, 554–565.

Potapov, P.V., Turubanova, S.A., Hansen, M.C., Adusei, B., Broich, M., Altstatt, A., et al. (2012). Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM C data. *Remote Sensing of Environment, 122*, 106–116.

Rabus, B., Eineder, M., Roth, A., & Balmer, R. (2003). The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. *Photogrammetric Engineering and Remote Sensing, 57*, 241–262.

Reed, B.C., Brown, J.F., Vanderzee, D., Loveland, T.R., Merchant, J.W., & Ohlen, D.O. (1994). Measuring phenological variability from satellite imagery. *Journal of Vegetation Science, 5*, 703–714.

Richards, J.A. (1999). *Remote sensing digital image analysis.* Berlin: Springer-Verlag, 240.

Rogan, J., Franklin, J., Stow, D., Miller, J., Roberts, D.A., & Woodcock, C. (2008). Mapping land cover modifications over large areas: A comparison of machine learning techniques. *Remote Sensing of Environment, 112*, 2272–2283.

Rosenblatt, F. (1957). The Perceptron—a perceiving and recognizing automaton. *Report 85-460-1.* Cornell Aeronautical Laboratory.

Rosenblatt, F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65*, 386–408.

Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M.C., et al. (2010). Web enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sensing of Environment, 114*, 35–49.

Roy, D.P., Qin, Y., Kovalskyy, V., Vermote, E.F., Ju, J., Egorov, A., et al. (2014). Conterminous United States demonstration and characterization of MODIS-based Landsat ETM+ atmospheric correction. *Remote Sensing of Environment, 140*, 433–449.

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., et al. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment, 145*, 154–172.

Savage, L.J. (1976). On rereading R. A. Fisher. *The Annals of Statistics, 4*(3), 441–500.

Settles, B. (2009). Active learning literature survey. *Computer Sciences Technical Report 1648.* University of Wisconsin-Madison.

Settles, B., & Craven, M. (2008). An analysis of active learning strategies for sequence labeling tasks. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1069–1078). ACL Press.

Stoker, J.M., Heidemann, H.K., Evans, G.A., & Greenlee, S.K. (2013). *A conceptual prototype for the next-generation national elevation dataset: U.S. Geological Survey Open-File Report 2013–1023.* (52 pp.).

SWBD (2005). Shuttle Radar Topography Mission water body data set. http://www2.jpl.nasa.gov/srtm/index.html (Digital Media (Last accessed May 30th 2014)

Thompson, C.A., Califf, M.E., & Mooney, R.J. (1999). Active learning for natural language parsing and information extraction. *Proceedings of the International Conference on Machine Learning (ICML)* (pp. 406–414). Morgan Kaufmann.

Tong, S., & Koller, D. (2000). Support vector machine active learning with applications to text classification. *Proceedings of the International Conference on Machine Learning (ICML)* (pp. 999–1006). Morgan Kaufmann.

Tuia, D., Pacifici, F., Kanevski, M., & Emery, W.J. (2009). Classification of very high spatial resolution imagery using mathematical morphology and support vector machines. *IEEE Transactions on Geoscience and Remote Sensing, 47*(11), 3866–3879.

Tuia, D., Volpi, M., Copa, L., Kanevski, M., & Munoz-Mari, J. (2011, June). A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing, 5*(3), 606–617.

Tür, G., Hakkani-Tür, D., & Schapire, R.E. (2005). Combining active and semisupervised learning for spoken language understanding. *Speech Communication, 45*(2), 171–186.

Vogelmann, J.E., Howard, S.M., Yang, L.M., Larson, C.R., Wylie, B.K., & Van Driel, N. (2001). Completion of the 1990s National Land Cover Data set for the conterminous United States from Landsat Thematic Mapper data and Ancillary data sources. *Photogrammetric Engineering and Remote Sensing, 67*, 650–652.

Wickham, J.D., Stehman, S.V., Gass, L., Dewitz, J., Fry, J.A., & Wade, T.G. (2013). Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment, 130*, 294–304.

Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., et al. (2008). Free access to Landsat imagery. *Science, 320*, 1011.

Yan, L., & Roy, D.P. (2015). Improved time series land cover classification by missing-observation-adaptive nonlinear dimensionality reduction. *Remote Sensing of Environment, 136*, 478–491.

Yan, R., Yang, J., & Hauptmann, A. (2003). Automatically labeling video data using multi-class active learning. *Proceedings of the International Conference on Computer Vision* (pp. 516–523). IEEE Press.

Yu, H. (2005). SVM selective sampling for ranking with application to data retrieval. *Proceedings of the International Conference on Knowledge Discovery and Data Mining (KDD)* (pp. 354–363). ACM Press.

Yu, B. -H., & Chi, K. -H. (2008). Support vector machine classification using training sets of small mixed pixels: An appropriateness assessment of IKONOS imagery. *Korean Journal of Remote Sensing, 24*(5), 507–515.

Zhuang, X., Engel, B.A., Lozanogarcia, D.F., Fernandez, R.N., & Johannsen, C.J. (1994). Optimization of training data required for neuro-classification. *International Journal of Remote Sensing, 15*, 3271–3277.