South Dakota State University
Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange

GSCE Faculty Publications | Geospatial Sciences Center of Excellence (GSCE)

8-2017

Using the 500 m MODIS Land Cover Product to Derive a Consistent Continental Scale 30 m Landsat Land Cover Classification

Hankui Zhang
South Dakota State University, hankui.zhang@sdstate.edu

David P. Roy
South Dakota State University, david.roy@sdstate.edu

Follow this and additional works at: http://openprairie.sdstate.edu/gsce_pubs

Part of the Geographic Information Sciences Commons, Physical and Environmental Geography Commons, and the Remote Sensing Commons

Recommended Citation
Zhang, Hankui and Roy, David P., "Using the 500 m MODIS Land Cover Product to Derive a Consistent Continental Scale 30 m Landsat Land Cover Classification" (2017). *GSCE Faculty Publications*. 30.
http://openprairie.sdstate.edu/gsce_pubs/30

This Article is brought to you for free and open access by the Geospatial Sciences Center of Excellence (GSCE) at Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. It has been accepted for inclusion in GSCE Faculty Publications by an authorized administrator of Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. For more information, please contact michael.biondo@sdstate.edu.

Remote Sensing of Environment 197 (2017) 15–34

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification

Hankui K. Zhang *, David P. Roy

Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD 57007, USA

ARTICLE INFO

Article history:
Received 21 February 2017
Received in revised form 23 April 2017
Accepted 20 May 2017
Available online 25 May 2017

Keywords:
Large area
Land cover
Classification
Landsat
MODIS
Monthly metrics

ABSTRACT

Classification is a fundamental process in remote sensing used to relate pixel values to land cover classes present on the surface. Over large areas land cover classification is challenging particularly due to the cost and difficulty of collecting representative training data that enable classifiers to be consistent and locally reliable. A novel methodology to classify large volume Landsat data using high quality training data derived from the 500 m MODIS land cover product is demonstrated and used to generate a 30 m land cover classification for all of North America between 20°N and 50°N. Publically available 30 m global monthly Web-enabled Landsat Data (GWELD) products generated from every available Landsat 7 ETM+ and Landsat 5 TM image for a three year period, that are defined aligned to the MODIS land products and are consistently pre-processed data (cloud-screened, saturation flagged, atmospherically corrected and normalized to nadir BRDF adjusted reflectance), were classified. The MODIS 500 m land cover product was filtered judiciously, using only good quality pixels that did not change land cover class in 2009, 2010 or 2011, followed by automated selection of spatially corresponding 30 m GWELD temporal metric values, to define a large training data set sampled across North America. The training data were sampled so that the class proportions were the same as the North America MODIS land cover product class proportions and corresponded to 1% of the 500 m and <0.005% of the 30 m pixels. Thirty nine GWELD temporal metrics for every 30 m North America pixel location were classified using (a) a single random forest, and (b) a locally adaptive method with a random forest classifier derived and applied locally and the classification results spatially mosaicked together. The land cover classification results appeared geographically plausible and at synoptic scale were similar to the MODIS land cover product. Detailed visual inspection revealed that the locally adaptive random forest classifications and associated classification confidences were generally more coherent than the single random forest classification results. The level of agreement between the 30 m classifications and the MODIS land cover product derived training data was assessed by bootstrapping the random forest implementation. The locally adaptive random forest classification had higher overall agreement (95.44%, 0.9443 kappa) than the single random forest classification (93.13%, 0.9195 kappa). The paper concludes with a discussion of future research including the potential for automated global land cover classification.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Satellite data are used to generate large area land cover maps needed to understand and census anthropogenic activity and the biogeographical and ecoclimatic diversity of the land surface (Loveland et al., 2000; Turner et al., 2007). Landsat data provide the longest terrestrial remote sensing record and have a long history for land cover mapping because of their moderate spatial resolution and near global coverage (Roy et al., 2014a; Wulder et al., 2016). The advent of free Landsat data combined with improving computational and data storage capabilities mean that large area Landsat land cover products are increasingly being generated.

Recently, 30 m global land cover products were generated using training data obtained by photo-interpretation of Google Earth imagery and supervised classification of single Landsat images (Gong et al., 2013) augmented by 250 m Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI time series (Yu et al., 2013; Chen et al., 2015). Although these global products provide useful information they typically have fewer land cover classes and lower accuracy than 30 m national land cover products derived from multi-temporal Landsat data, such as, for example, the 5-year United States Geological Survey (USGS) National Land Cover Database (NLCD) that has 16 classes and 82% overall accuracy (Homer et al., 2015; Wickham et al., 2017), or the annual United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) that has 110 classes and 84% overall accuracy (Boryan et al., 2011; Johnson, 2013).

* Corresponding author.
E-mail addresses: hankui.zhang@sdstate.edu (H.K. Zhang), david.roy@sdstate.edu (D.P. Roy).

http://dx.doi.org/10.1016/j.rse.2017.05.024
0034-4257/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

The current state of the practice for large area multi-temporal land cover classification is to derive metrics from the time series and then classify the metrics bands with a supervised (i.e., training data dependent) non-parametric classification approach (Yan and Roy, 2015; Gómez et al., 2016). The random forest classifier (Breiman, 2001) is a non-parametric classifier commonly used for land cover mapping (Lawrence and Moran, 2015; Inglada et al., 2015; Wessels et al., 2016; Belgiu and Drăguţ, 2016; Gong et al., 2013; Hermosilla et al., 2017). Random forests are an ensemble form of decision tree classification. Unlike decision tree classifiers, each tree is grown using randomly selected predictor variables, and not just using a random subset of the training data, to reduce the likelihood of over-fitting the predictor variables to the training data (Breiman, 2001). In addition, the random forest classifier may be less sensitive to noise (in the training data and/or satellite data classified) and may be more efficient than other commonly used non-parametric classifiers such as support vector machines (Pelletier et al., 2016).

Supervised classification approaches are not automated due to their reliance on training data collection with the majority of the classification effort expended on training data collection and refinement (Huang et al., 2008; Townshend et al., 2012; Gong et al., 2013; Egorov et al., 2015). Training data should be selected to capture all relevant spectral heterogeneity within and among classes (Foody and Mathur, 2006) and ideally should be collected in a way that satisfies probability sampling design criteria (Stehman, 2001; Boschetti et al., 2016). Over large areas the optimal training size and distribution is usually unknown and is dependent on the satellite data and the classifier used. For non-parametric classifiers the naturally occurring class distribution, i.e., a proportional distribution among the classes related to the proportion that they occur in reality, or an equal balance of the training data among the classes, provide reasonable classification accuracies provided that there are sufficient training data (Weiss and Provost, 2003; Colditz, 2015). What constitutes a sufficient amount of training data is hard to define *a priori* although classification accuracy generally increases with training set size (Rogan et al., 2008; Yan and Roy, 2015). The application of training data derived class signatures to classify other locations or times becomes less appropriate the further away in space and/or time that they are applied (Henderson, 1976; Woodcock et al., 2001). For large area classification researchers have attempted to overcome this issue by independent classification of geographic strata, for example, latitudinal strata (DeFries and Townshend, 1994), ecologically defined strata (Loveland et al., 1991; Homer et al., 2004; Schneider et al., 2010), or individual images (Gong et al., 2013; Homer et al., 2015) although it is recognized that stratum specific training samples may not be available if the strata are small and classification inconsistencies may occur along strata edges.

In the last decade the use of existing land cover maps as a source of training data has been demonstrated (Knorn et al., 2009; Xian et al., 2009; Sexton et al., 2013; Jia et al., 2014; Radoux et al., 2014; Wessels et al., 2016). This is advantageous as it (i) enables the classification to be undertaken in an automated manner without the need for interactive and manual training data collection and refinement, (ii) provides a potentially large and geographically distributed training data set, and (iii) enables the satellite data to be classified with the same legend as the existing land cover map. Care must be taken to appropriately filter the land cover map to ensure that only reliably defined training data are extracted and, as in this study, to accommodate any spatial resolution or temporal reporting differences between the land cover map and the satellite data that are to be classified.

The objective of this study is to demonstrate an automated continental scale 30 m Landsat land cover classification of the recently available global Web-enabled Landsat Data (GWELD) products using the MODIS 500 m land cover product (Friedl et al., 2010) as a source of training data. GWELD products generated from every available Landsat 7 ETM + and Landsat 5 TM image for a three year period for all of North America between 20°N and 50°N were classified. Over such a large area the satellite data must be pre-processed consistently to ensure that training data are broadly applicable (Gray and Song, 2013). The GWELD products are defined in the same projection as the MODIS land cover product in 30 m tiles and provide consistently pre-processed data, i.e., cloud-screened, saturation flagged, atmospherically corrected and normalized to nadir BRDF adjusted reflectance (NBAR) (Roy et al., 2010a, 2016a). The MODIS 500 m land cover product was filtered judiciously, using only good quality pixels that did not change land cover class in 2009, 2010 or 2011, followed by automated selection of spatially corresponding 30 m GWELD metrics values, to define a large training data set. Thirty nine GWELD 30 m temporal metrics were classified. The metrics were extracted for each 30 m pixel location considering three years (2009, 2010 and 2011) of GWELD data together. Two random forest classifications were undertaken using the MODIS land cover product legend excluding the urban and built-up class. First, a single random forest was applied to all the North America GWELD metric data. Second, a locally adaptive random forest was applied to individual GWELD metric tiles and the tile classification results were mosaicked together. The locally adaptive random forest classification used the single random forest classification training data as default but with updated training samples available from each tile locality. The training data were sampled so that the relative frequency of the land cover classes were proportional to their occurrence across North America while ensuring that all the classes were represented. The level of agreement between the 30 m classifications and the MODIS land cover product derived training data was assessed by bootstrapping the random forest implementation. In addition, maps of the classification confidence, defined for each pixel as the proportion of times over the different random forest trees that the pixel was classified as the majority class, were assessed. The classification results were compared with the MODIS land cover product to gain insights into the scale differences between the 30 m and 500 m classifications.

2. Data and study area

2.1. Three years of MODIS land cover products

The global Collection 5 annual 500 m MODIS land cover product (MCD12Q1) (Friedl et al., 2010) for 2009, 2010 and 2011 were used to define land cover training class labels. The MCD12Q1 International Geosphere-Biosphere Program (IGBP) classification scheme, which classifies each 500 m pixel into one of 17 classes (Table 1) and has a reported 75% overall land cover classification accuracy (Friedl et al., 2010) was used. In addition, the MCD12Q1 500 m classification confidence (Land_Cover_Type_1_Assessment) and quality assessment (Land_Cover_Type_QC) data layers were used to help select only good quality and high confidence training class labels. The MCD12Q1 product is defined in the standard MODIS 10° × 10° MODIS land product tile system in the equal area sinusoidal projection (Wolfe et al., 1998).

2.2. Three years of global Web-enabled Landsat Data

Monthly 30 m global Web-enabled Landsat Data (GWELD) version 3.0 products publically available from http://globalweld.cr.usgs.gov/collections/ (and see http://go.nasa.gov/2kLcKto for visualizations) were classified to generate 30 m land cover maps defined with the MODIS IGBP land cover classification scheme (Table 1). The GWELD products have heritage from the Web Enabled Landsat Data (WELD) products that were generated for ten years over the continuous United States (CONUS) and Alaska (Roy et al., 2010a). The WELD products have been used to make CONUS spatially explicit 30 m maps of percent tree cover, bare ground and other vegetation and their change (Hansen et al., 2011, 2014), surface water and permanent snow (Egorov et al., 2015), burned areas (Boschetti et al., 2015), and crop fields (Yan and Roy, 2016). In addition to providing global coverage, the GWELD products have several algorithm improvements over the WELD products.

Table 1
The 17 IGBP MODIS MCD12Q1 land cover classes (Friedl et al., 2010) and the percentage of the study area that they cover as defined by the 2010 MCD12Q1 product.

Class	Name	CONUS area %	Class	Name	CONUS area %
0	Water	3.41	9	Savannas	0.46
1	Evergreen needleleaf forest	6.27	10	Grasslands	23.60
2	Evergreen broadleaf forest	0.73	11	Permanent wetlands	0.64
3	Deciduous needleleaf forest	0.05	12	Croplands	14.31
4	Deciduous broadleaf forest	4.81	13	Urban and built-up[a]	1.23
5	Mixed forest	12.50	14	Cropland/natural vegetation mosaic	10.44
6	Closed shrublands	0.38	15	Snow and ice	0.02
7	Open shrublands	12.53	16	Barren or sparsely vegetated	1.24
8	Woody savannas	7.38			

[a] Class 13 (urban and built-up) was not used in the GWELD 30 m classification.

First, they are generated from every available Landsat 5 Thematic Mapper (TM) and 7 Enhanced Thematic Mapper plus (ETM+) Level 1 T image held in the United States Landsat archive at the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center. Second, the reflective wavelength bands are atmospherically corrected using the established Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) that uses a radiative transfer code with aerosol characterization derived independently for each Landsat acquisition and using external water vapor and ozone characterizations (Masek et al., 2006). Third, the reflective wavelength bands are corrected to nadir BRDF-adjusted surface reflectance (NBAR) using a c-factor BRDF normalization method and a fixed set of MODIS BRDF spectral model parameters (Roy et al., 2016a) with a solar zenith definition modeled to reflect the Landsat 5 and 7 solar zenith that was observed globally in 2011 (Zhang et al., 2016). Fourth, the GWELD products are stored in tiles that are nested within the 10° × 10° MODIS land product tiles (Fig. 1) so it is straightforward to compare the 30 m GWELD products with any of the standard gridded MODIS land products (Justice et al., 2002).

Each GWELD tile is composed of 5295 × 5295 30 m pixels. There are 7 × 7 GWELD tiles within each MODIS land tile (Fig. 1). The GWELD tile locations are reflected in the filename, designated as hh⟨*xx*⟩vv⟨*yy*⟩.h⟨*x*⟩v⟨*y*⟩, where *xx* and *yy* are the standard two digit horizontal (0 to 35) and vertical (0 to 17) MODIS land tile coordinates (Wolfe et al., 1998), and *x* and *y* are one digit horizontal (0 to 6) and vertical (0 to 6) GWELD tile coordinates. Each GWELD tile pixel defines for each month or year the "best" Landsat 5 TM or Landsat 7 ETM+ observation data available at the 30 m pixel tile location. The compositing method is based on the method described in Roy et al. (2010a). The information stored at each GWELD 30 m tile pixel location include: the atmospherically corrected NBAR for each Landsat reflective wavelength band and the derived NDVI, the top of atmosphere brightness temperature, the date each composited pixel was acquired on, the per-band radiometric saturation status, two cloud mask values that were derived from the heritage Landsat project automatic cloud cover assessment algorithm (ACCA) (Irish et al., 2006) and a decision tree cloud mask algorithm (Roy et al., 2010a), the number of acquisitions considered in the compositing period, the sensor (Landsat 5 or 7), the observed solar and sensor angles, and the solar zenith that the pixel was normalized to, and also an index that can be used to define the L1T image and the L1T pixel row and column location (Roy et al., 2014b).

Monthly GWELD products over the CONUS for three climate years (i.e., 36 months) from December 2009 to November 2011 were used. Annual GWELD products are also available but were not used in this study as they were not found to improve the land cover classification accuracy. Only the GWELD Landsat TM and ETM+ atmospherically corrected NBAR for bands 2 (green, 0.53–0.61 μm), 3 (red: 0.63–0.69 μm), 4 (near-infrared: 0.76–0.90 μm for TM and 0.77–0.90 μm for ETM+), 5 (middle-infrared: 1.55–1.75 μm), and 7 (middle-infrared: 2.08–2.35 μm for TM and 2.09–2.35 μm for ETM+) were used. The

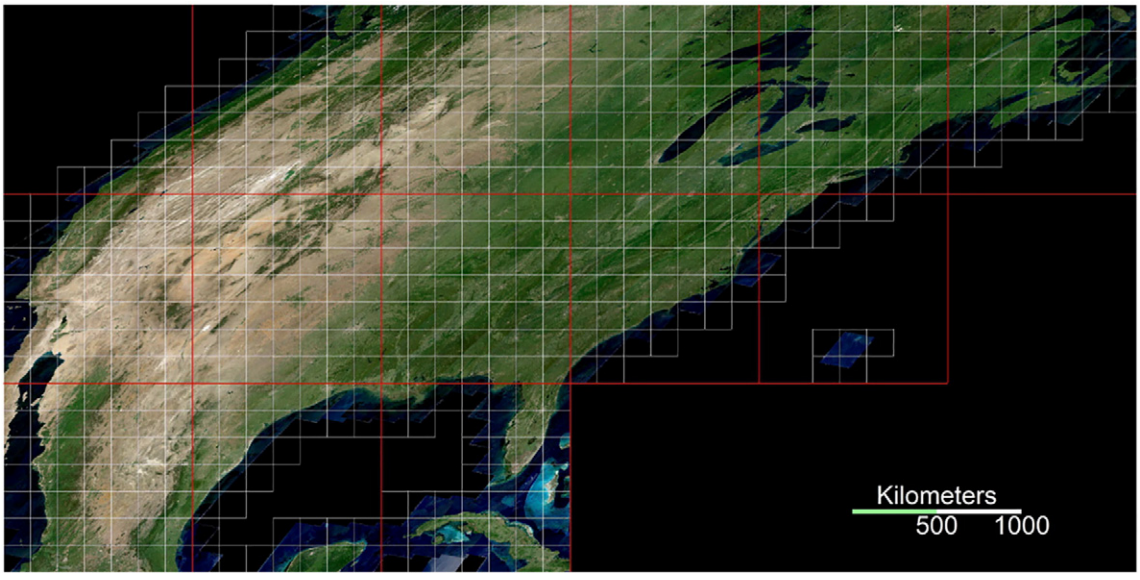

Fig. 1. Illustration of the 20°N and 50°N CONUS study area, composed of 561 GWELD 5295 × 5295 30 m pixel tile boundaries (white) that are spatially nested within 14 standard MODIS land 10° × 10° tile boundaries (red) defined in the sinusoidal equal area projection. Note that of the 561 GWELD tiles a total of 511 were classified (as 50 were labelled by the MODIS land cover product quality assessment layer as shallow ocean, moderate/continental ocean, or deep ocean and were not considered). The background shows for geographic context the year 2010 version 3.0 GWELD true color NBAR product.

shortest wavelength Landsat TM and ETM+ band 1 (blue: 0.45–0.52 μm) was not used because it is very sensitive to atmospheric scattering and is not reliably atmospherically corrected (Ju et al., 2012; Roy et al., 2014b; Claverie et al., 2015). All 30 m GWELD pixel values flagged as cloudy in both the cloud masks, or as saturated, were discarded from the analysis.

2.3. Study area

The study area is defined by 14 North America 10° × 10° MODIS land tiles (Fig. 1, red) located between 20°N and 50°N that encompass all of the CONUS and parts of northern Mexico, the Caribbean, and southern Canada. Within the 14 MODIS tiles there were a total of 561 GWELD tiles (Fig. 1, white). A total of 29,292 Landsat 5 TM and 26,686 Landsat 7 ETM+ Level 1 T images were used to generate the three years of GWELD data for these tiles. Fifty GWELD tiles were only shallow ocean, moderate/continental ocean, or deep ocean (as labelled in the MCD12Q1 quality assessment layer) and so were removed. This reduced the number of GWELD tiles that were classified from 561 to 511 tiles.

2.4. National Land Cover Database (NLCD) urban layer

It is well established that urban areas are difficult to classify reliably because they encompass a variety of land cover types and land uses that even at Landsat 30 m scale are often mixed spatially (Herold et al., 2003; Small, 2005; Lu et al., 2008; Griffiths et al., 2010). In this study GWELD land cover classification was undertaken without using the MODIS land cover product urban and built-up class (Table 1). Instead, a static 30 m urban mask derived from the 2011 CONUS National Land Cover Database (NLCD) (Homer et al., 2015) was superimposed over the GWELD classifications. The 2011 NLCD is available for the CONUS and includes four urban classes: Developed low, Developed medium, and Developed high intensity, and an Open Space class (Homer et al., 2004). These classes were derived from a percent imperviousness map generated using regression tree techniques applied to Landsat data and ancillary data including NOAA Defense Meteorological Satellite Program (DMSP) night light data, road vector data, and USGS National Elevation Dataset (NED) digital elevation data (Yang et al., 2003; Homer et al., 2015).

The static CONUS 30 m urban mask was defined using the NLCD 2011 Developed low, Developed medium, and Developed high intensity classes. The NLCD 2011 Open Space class, that defines vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes (Homer et al., 2004), was not used as its definition is inconsistent with the MODIS land cover product urban and built up class definition (Schneider et al., 2010). The 2011 NLCD has been validated with an overall classification accuracy of 82% and with urban class user's and producer's accuracies of 84% and 80% respectively (Wickham et al., 2017). The 2011 NLCD data are defined in the Albers projection and so were re-projected into the MODIS sinusoidal projection by nearest neighbor resampling to preserve the classification label values.

3. Method

3.1. WELD monthly metric generation

The supervised non-parametric classification of temporal metrics derived from Landsat time series is a widely used approach to generate large area land cover classifications (Hansen et al., 2014; Yan and Roy, 2015; Wessels et al., 2016; Gomez et al. 2016). Temporal metrics, such as the median value, are insensitive to phenological differences and missing data. Temporal metrics do not explicitly capture the timing but rather the amplitude of the reflectance variation and so are insensitive to phenological differences (DeFries et al., 1995; Friedl et al., 2010) where time series may exhibit different phenological variation at different locations for the same land cover class (Zhang et al., 2006). Temporal metrics are robust to missing data which is important because Landsat time series have gaps due to cloud cover (Kovalskyy and Roy, 2013), variable Landsat acquisition frequency (Wulder et al., 2016), and sensor issues (Markham et al., 2004).

The metrics were similar to those used previously to classify 30 m percent tree cover, bare ground and other vegetation for all the CONUS using WELD data (Hansen et al., 2011). Specifically, the 20th, 50th (i.e., median) and 80th percentiles of Landsat NBAR bands 2, 3, 4, 5, 7, and of eight normalized NBAR band ratios $4-3/4+3$ (i.e. NDVI), $5-2/5+2$, $5-3/5+3$, $5-4/5+4$, $7-2/7+2$, $7-3/7+3$, $7-4/7+4$, and $7-5/7+5$, were used. The 20th and 80th percentiles were used, rather than minimum and maximum values, to reduce sensitivity to shadows and residual cloud and atmospheric contamination effects. This provided a total of 39 metrics for each 30 m GWELD pixel location. The metrics were extracted considering only April to October because of persistent cloud and snow in the CONUS winter (Ju and Roy, 2008; Hansen et al., 2011). Thus, a maximum of 21 possible (seven months of 2009, 2010, 2011) unsaturated and cloud-free monthly observations were used to derive the metrics without consideration of the acquisition year, which is a common approach when inter-annual variation is limited and to ensure more temporal observations are available (Zhu and Woodcock, 2014; Schmidt et al., 2016). If there were less than five unsaturated or cloud-free observations then the GWELD metrics were considered invalid as there were too few to reliably define the 20th, 50th and 80th percentiles.

3.2. Study area 30 m land cover training pool derivation

Care was taken to ensure that only reliable and representative training data were extracted as the training data influence directly the classification accuracy (Foody and Mathur, 2004). A large set of 30 m training data composed of MODIS land cover class labels (Table 1) and associated 39 GWELD metrics were extracted systematically across the study area. For brevity we refer to this as the training pool.

The training pool generation first required the selection of suitable MODIS 500 m land cover pixels. Only MCD12Q1 pixels that were classified consistently as the same land cover class over the three years (2009 to 2011) and that always had classification confidence (Land_Cover_Type_1_Assessment) >50% and quality assessment (Land_Cover_Type_QC) set as "good quality" were considered. Pixels classified by MCD12Q1 as Urban and built-up, or labelled the MCD12Q1 quality assessment layer as shallow ocean, moderate/continental ocean, or deep ocean, were not considered. For all the classes, except the deciduous needleleaf forest class, a spatial filter was applied so that only the MCD12Q1 pixel locations that had the same land cover class in the surrounding eight 500 m pixels were retained. This is similar to previous approaches (Blanco et al., 2013; Colditz et al., 2012) and was implemented to help reduce spatial differences between the 500 m MCD12Q1 and 30 m GWELD data, in particular, 500 m pixel edge effects where the underlying land cover may change across 500 m pixel boundaries, and also to reduce the impact of the 50 m 1σ MODIS geolocation error and variable across-track MODIS spatial resolution (Wolfe et al., 2002; Campagnolo et al., 2016). The deciduous needleleaf forest class was not subject to the spatial filtering as no good quality consistent deciduous needleleaf forest 500 m pixels remained after the spatial filtering. We note that this class is present in North America (Friedl et al., 2000) but only with a sparse geographic distribution (Table 1). Fig. 2 illustrates the selection results for a single 2010 MCD12Q1 tile (left) and the reduced number of 500 m pixels after the filtering. In this illustrated example 13.71% of the 500 m pixels (Fig. 2 left) remain after the filtering (Fig. 2, right).

Within each of the filtered MCD12Q1 500 m product pixels a single 30 m GWELD pixel location was selected. This is complicated because the spatial arrangement of land cover may be quite different at 500 m and 30 m. This has not been studied for the CONUS but, for example, Roy and Kumar (2017) reported that only about 5% of 1 km MODIS

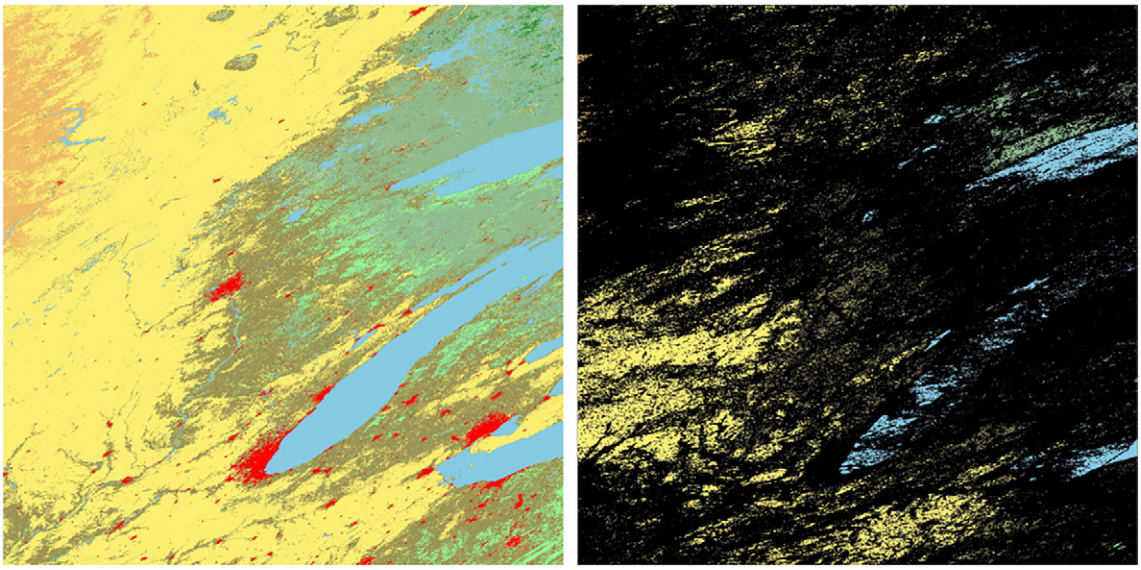

Fig. 2. Illustration of the MCD12Q1 500 m filtering to select the class labels and where the training pool data were defined. *Left:* example MCD12Q1 2010 500 m land cover product (colors correspond to different land cover classes, Table 1) over MODIS land tile h11v04 that covers approximately 1200 × 1200 km (109.0855° to 78.1497°W, 39.7728° to 50.0921°N) and includes Minneapolis, Chicago and Detroit. *Right:* the filtered 500 m MCD12Q1 pixels that were classified consistently as the same land cover class over the three years (2009 to 2011) and that always had classification confidence (Land_Cover_Type_1_Assessment) >50% and quality assessment (Land_Cover_Type_QC) set as "good quality". Pixels classified as Urban and built-up (red in left image) were not considered.

pixels over the Brazilian Tropical Moist Forest Biome (4 million km^2) contained homogeneous land cover mapped at 30 m. In this study the following selection process was implemented and is based on the method described in Roy et al. (2016a). The GWELD "metric centroid" of all the 30 m Landsat pixel values falling within the 500 m MODIS pixel was defined as:

$$\boldsymbol{m}_c = \sum_{i=1}^{n} \boldsymbol{m}_i / n \quad (1)$$

where $\boldsymbol{m}_c$ (the metric centroid) is a vector of 39 metric average values $(m_c^1, m_c^2, \ldots, m_c^{39})$ defined from the n 30 m pixels that fall within the 500 m pixel, and $\boldsymbol{m}_i$ is a vector of the 39 GWELD metric values $(m_i^1, m_i^2, \ldots, m_i^{39})$ for 30 m pixel i. The 30 m pixel that was selected was the one that minimized:

$$\Delta_i = |\boldsymbol{m_i} - \boldsymbol{m_c}| \quad (2)$$

where Δ_i is the absolute distance between the metric centroid vector $\boldsymbol{m}_c = [m_c^1, m_c^2, \ldots, m_c^{39}]$ and the metric vector at pixel i ($\boldsymbol{m}_i = [m_i^1, m_i^2, \ldots, m_i^{39}]$). If several 30 m pixels had the same Δ value then one pixel was selected at random. The underlying assumption of Eqs. (1) and (2) is that the majority of the 30 m pixels have the same land cover type as the 500 m pixel. Consequently, only MODIS 500 m pixel locations that were >75% covered by 30 m pixels with valid GWELD metrics (i.e., 30 m pixel locations where there were five or more unsaturated cloud-free observations over the three years) were considered.

The above procedure was used to generate a training pool defined by a table with 40 columns (the single MCD12Q1 500 m land cover class label and the corresponding 39 GWELD metrics at each selected 30 m location) and several million rows (different 500 m pixels across the study area). The sampling, and so the population of the table rows, was undertaken systematically from the north west to the south east across the study area.

3.3. Random forest land cover classification

3.3.1. Overview

Two types of random forest classifications were undertaken. First, a single random forest was derived and used to classify all the 511 GWELD tiles. Second, a locally adaptive random forest was derived for each GWELD tile and the resulting 511 tile classification results were spatially mosaicked together. Both used sub-sets of the training pool sampled so that the relative class frequency in the training data was proportional to the study area MCD12Q1 class frequency while ensuring that all 16 classes (all but the Urban and built-up class, Table 1) were included.

The classifications were undertaken using the R RANDOMFOREST package (http://www.r-project.org/) with default parameter settings (Liaw and Wiener, 2002). A total of 500 trees were grown with each tree considering 63.2% of the training data selected at random without replacement and considering six randomly selected GWELD metrics per partition in the tree. Six randomly selected GWELD metrics per partition were used as this number is approximately the square root of the number of available predictor variables (the 39 GWELD metrics). Each tree was used to independently classify the GWELD metrics for each 30 m pixel and the final land cover class was assigned in the conventional way as the majority class over the 500 classifications. The resulting classifications were composed of 16 classes with no urban built-up class (Table 1).

Post-classification, the static NLCD-based CONUS 30 m urban mask was used to provide a final CONUS classification including an urban definition. Outside the CONUS (approximately 122 of the 511 GWELD study area tiles) there is no NLCD product. Thus, all CONUS 30 m pixel locations that were labelled in the NLCD as Developed high, medium or low urban density were labelled in the final GWLED classification as Urban and built-up.

3.3.2. Single random forest 30 m land cover classification

A single random forest was applied to all the 511 GWELD tiles. The number of selected training data samples for each class was defined by the parameter p as:

$$n_i^* = \frac{p}{100}\, n_i^{MODIS} \quad (3)$$

where n_i^* is the number of samples for land cover class i selected from the training pool, p is the percentage (>0 to 100) of the North American 500 m land pixels considered, and n_i^{MODIS} is the total number of study area MCD12Q1 2010 500 m pixels for class i. Different p values were considered and the derived n_i^* values examined to ensure that each land cover class had sufficient training data selected from the training

pool. Due to the training pool derivation (Section 3.2) for a given class i the number of class samples in the training pool (n_i) may be smaller than n_i^* in which case n_i^* was set as n_i.

The selection of the training data from the training pool was not undertaken randomly but rather in a geographically systematic manner. Recall that the training pool was defined as a table where the row order was populated systematically across the study area from the north west to the south east. A separate table for each land cover class i was extracted and then a selection was undertaken from each by extracting every n_i^*/n_i row where n_i^* is defined as Eq. (3), n_i is the number of class samples in the training pool, and n_i^*/n_i was rounded to the nearest integer and is ≥1. This helped to ensure that the training data for each land cover class were distributed across the study area which may not occur using a simple random sampling scheme.

3.3.3. Locally adaptive random forest 30 m land cover classification

Locally adaptive random forest classifications were undertaken in an attempt to reduce signature extension issues. It is well established that the application of training data class signatures to classify other locations or times becomes less appropriate if the land cover spectral signatures are different further away in space or time (Henderson, 1976; Minter, 1978; Woodcock et al., 2001). For example, consider the MCD12Q1 deciduous broadleaf forest class that has a CONUS range extending from Maine to Florida. The deciduous broadleaf forest training data collected over this range will include quite different forest species, forest structural compositions, understory vegetation, and soils. These signature extension issues may be confounded by noisy data and residual data pre-processing errors (Gray and Song, 2013) although we expect this to be reduced by the GWELD processing and the use of metrics.

The local areas were defined by GWELD tiles as they have approximately the same area (159 × 159 km) as Landsat images (170 km × 180 km) that are used as a spatial unit for national (Homer et al., 2015) and global land cover mapping (Gong et al., 2013). The class proportions in each tile defined by MCD12Q1 were not considered because of the spatial resolution difference between the 500 m and 30 m pixels — tiles with a small number of MCD12Q1 500 m classes may contain more classes at 30 m (this is illustrated in the results). In addition, the training pool data were not available everywhere due to the filtering used to generate them (e.g., Fig. 2). For each GWELD tile a random forest was generated using, as a default, the single random forest training data (Section 3.3.2). Local training pool data may exist for the tile that were not selected to generate the single random forest and these were preferentially used in the local classification. Specifically, the single random forest training data were replaced with local training pool data while ensuring that for each class i there were n_i^* training data samples. In this way, training samples for all the land cover classes were always used, including classes that were not present locally. To ensure classification consistency across the GWELD tile boundaries the training data were sampled locally from the training pool for 3 × 3 adjacent tiles to build a random forest that was then used to classify the central tile. Finally, the 511 independently derived central tile classifications were spatially mosaicked together.

3.4. Random forest land cover assessment

3.4.1. Quantitative accuracy assessment against the training pool data

The selected training pool data were considered as "truth" for the classification and for the following classification accuracy assessment. The selected MCD12Q1 class labels of the training pool data (classified consistently as the same MODIS land cover class over the three years with classification confidence >50% and quality assessment set as "good quality") were considered as without error. The classification accuracy was assessed quantitatively by bootstrapping the random forest implementation (Breiman, 2001) to derive confusion matrix based accuracy metrics. For brevity we sometimes refer to the metric results in terms of "accuracy" but strictly they quantify the *agreement* between the classifications and the MCD12Q1 training pool data. This is discussed further at the end of this paper.

For the single random forest classification, after each of the 500 trees was generated using 63.2% of the training data, the remaining 'out-of-bag' (OOB) 32.8% sample was classified with the tree and the classified out of bag (OOB) results stored. Thus, for 500 trees every training data sample was typically considered as an OOB sample $n = 164$ (0.328 × 500) times. For each training sample the majority class over the n OOB classifications was the classification result. This result, and the corresponding training pixel MCD12Q1 class label, was used to populate a two-way confusion matrix. The locally adaptive random forest classification accuracy was assessed in the same way except that only the OOB samples located within the 511 central GWELD tile were retained. In this way, one confusion matrix was derived using all 511 central tile OOB samples. Conventional accuracy statistics, i.e., percent correct (0–100%), kappa (0–1), and land cover class user's and producer's accuracies (0–100%) (Foody, 2002), were derived from the two confusion matrices.

The entire training pool was not used for the accuracy assessment because, as described in the results (Section 4.1), the training pool selection criteria meant that several of the classes had relatively small numbers of training pool pixels and so they were all used to generate the classifications. Only the accuracy of the classes in the random forest classifications could be considered due to the bootstrapping and so the accuracy of the NLCD derived urban mask was not measured.

3.4.2. Classification quality confidence maps

Spatially explicit 30 m maps of classification confidence were derived to provide insights into the classification performance. They do not measure accuracy but rather provide an indication of the classification quality. The confidence was defined for each pixel as the proportion of times over the 500 trees that the pixel was classified as the majority class (McIver and Friedl, 2001). The maximum possible confidence was 1.0 (all 500 trees classify a pixel as the same class) and the minimum was 0.0625 (= 1/16, i.e., when all the 16 classes are evenly distributed among the 500 trees). Other confidence measures, such as the second most common class or the number of unique classes that the pixel was classified as over the different trees (Dieye et al., 2011; Friedl et al., 2010) were implemented but due to paper length restrictions their results are not included in this study.

3.4.3. Comparison with the MCD12Q1 land cover classification

To complement the above assessments the 2010 MCD12Q1 500 m land cover product was compared to the 30 m random forest classifications except for locations labelled in the MCD12Q1 quality assessment layer as shallow ocean, moderate/continental ocean, or deep ocean. As the MCD12Q1 product includes the Urban and built-up class the 30 m NLCD urban mask was considered to ensure an unbiased comparison. Consequently, the following comparison was restricted to the CONUS where the NLCD is defined. The comparison does not assess the accuracy or quality of the random forest classifications but provides insights into the relative over- or under-estimation of the class proportions between the two scales of classification.

The 2010 MCD12Q1 500 m land cover pixels were resampled to 30 m and compared on a pixel by pixel basis with the 30 m random forest land cover classifications. For each of the 17 land cover classes (Table 1) the proportion of the resampled 30 m MCD12Q1 pixels classified as another (or the same) class in the 30 m Landsat random forest classification was derived considering all the CONUS pixels as:

$$p_a^b = \frac{\sum_{i,j \in \text{CONUS}} \left((MCD12(i,j) == a) \&\& (Landsat(i,j) == b) \right)}{\sum_{i,j \in \text{CONUS}} (MCD12(i,j) == a)} \quad (4)$$

where p_a^b is the proportion of the resampled 30 m MCD12Q1 class a pixels classified as class b in the 30 m random forest land cover classification, and $MCD12(i,j)$ and $Landsat(i,j)$ denote the MCD12Q1 and the

Landsat random forest classification values at pixel (i,j) respectively. This was undertaken for all class combinations, i.e., for class numbers $a \in \{0 \ldots 16\}$ and $b \in \{0 \ldots 16\}$. The 500 m MCD12Q1 pixels may be classified as several other classes in the 30 m random forest classifications and this is captured for all the CONUS by the range of $p_a^{b \in \{0 \ldots 16\}! = a}$ values for each class a.

4. Results

4.1. Training pool data and sub-sampling analysis

The training pool was defined by 4,182,823 different 500 m pixel locations with MCD12Q1 land cover class labels and 39 associated GWELD 30 m metrics. This corresponded to 8.44% of the total number of the study area MCD12Q1 2010 500 m land pixels. The training pool data were distributed across the study area and only 20 of the 511 GWELD tiles had no training pool data of which 16 were coastal tiles with only small land portions. However, when considering 3 × 3 neighboring GWELD tiles there was only one tile with no training pool data (over Bermuda in the Atlantic Ocean). Table 2 summarizes the total number of samples by land cover class in the training pool and their relative percentages. Evidently, due to the training pool derivation (Section 3.2), the relative percentage of training pool class samples is different to the study area MCD12Q1 2010 class samples that are also summarized.

Table 3 summarizes the number of training samples when $p = 0.2\%$, 0.4%, 0.6%, 0.8%, 1.0%, 1.5%, 2.0% of the North American MODIS land pixels are extracted balanced to the study area MCD12Q1 class proportions as Eq. (3). The grassland (class 10) and the snow and ice (class 15) classes have the greatest and smallest number of selected training data reflecting their relatively common and rare occurrence in the study area MCD12Q1 2010 classification (Table 2), respectively. In general, the number of class training samples increases with p. For some classes, in particular snow and ice (class 15), but also closed shrublands (class 6), woody savannas (class 8), and savannas (class 9), there were insufficient training samples in the training pool relative to the frequency of the other classes to enable balancing to the study area MCD12Q1 class proportions.

Table 4 summarizes the average percentage of study area land cover class training data samples (Table 3) that can be replaced by local training pool data over the 511 land 3 × 3 GWELD tile neighborhoods for different p values. For each class the average percentage replaced reduces with greater p. This is reasonable as greater p provides more training samples in the study area MCD12Q1 class portion balanced training data set (Table 3) and so fewer local samples remain in the training pool. For woody savanna (class 8) there were never any (0% on average, Table 4) additionally available samples locally.

The selection of a suitable p value for training data sampling from the training pool is an ill-posed optimization problem. This is illustrated in Fig. 3 which shows the impact of changing the p value ($p = 0.2\%$, 0.4%, 0.6%, 0.8%, 1.0%, 1.5%, 2.0% as Tables 3 and 4) on the single (black dotted lines) and locally adaptive (black solid lines) random forest classification accuracies for three example GWELD tiles. Strictly, the plotted accuracy results quantify the bootstrapped level of agreement between the classifications and the MCD12Q1 training pool data but for brevity we refer to the agreement as accuracy in this section.

For large p (not illustrated in Fig. 3) the single and the locally adaptive classification accuracies converge. The single random forest classification accuracy generally increases with p as more training data are extracted from the training pool (Table 3) and so there is an increased likelihood of capturing spectral heterogeneity within and among classes. Conversely, larger p values decrease the percentage of local training samples relative to the total used (shown by red lines) and so the locally adaptive classifications become less locally representative. Thus, typically, the locally adaptive classification accuracies (black solid lines) decrease with greater p. The exception is the Minnesota example (Fig. 3 right) that exhibits increasing locally adaptive classification accuracy

Table 2
Number of study area (Fig. 1) training pool data samples (n_i) for each land cover class and their relative percentages ($rp_i = 100\, n_i / \sum_i n_i$). For comparison the relative percentages ($rp_i^{MODIS} = 100 \frac{n_i^{MODIS}}{\sum_i n_i^{MODIS}}$) of the number of MCD12Q1 study area 500 m land cover class pixels (n_i^{MODIS}) are shown. Results are for all land cover classes (Table 1) except the Urban and built-up class.

	Land cover class														Row sum
	0. Water	1. Evergreen needleleaf forest	2. Evergreen broadleaf forest	3. Deciduous needleleaf forest	4. Deciduous broadleaf forest	5. Mixed forest	6. Closed shrublands	7. Open shrublands	8. Woody savannas	9. Savannas	10. Grasslands	11. Permanent wetlands	12. Croplands	14. Cropland/natural vegetation mosaic	15. Snow and ice	16. Barren or sparsely vegetated	
n_i (rp_i)	585,392 (13.995)	398,419 (9.525)	53,162 (1.271)	1642 (0.039)	63,731 (1.524)	588,467 (14.069)	2672 (0.064)	276,137 (6.602)	5690 (0.136)	1449 (0.035)	1,045,398 (24.993)	16,940 (0.405)	974,082 (23.288)	130,486 (3.120)	193 (0.005)	38,963 (0.932)	4,182,823 (100)
n_i^{MODIS} (rp_i^{MODIS})	1,712,125 (3.455)	3,148,156 (6.353)	364,025 (0.735)	22,585 (0.046)	2,413,668 (4.870)	6,273,101 (12.658)	189,267 (0.382)	6,286,373 (12.685)	3,702,503 (7.471)	230,177 (0.464)	11,841,624 (23.895)	320,098 (0.646)	7,179,896 (14.488)	5,238,514 (10.571)	10,495 (0.021)	624,665 (1.260)	49,557,272 (100)

Table 3
Number of study area (Fig. 1) training data samples (n_i^*) for each land cover class considering different percentages (p) of the study area 500 m pixels and balancing the class proportions according to the MCD12Q1 class proportions as Eq. (3).

p (%)	Land cover class																Row sum
	0. Water	1. Evergreen needleleaf forest	2. Evergreen broadleaf forest	3. Deciduous needleleaf forest	4. Deciduous broadleaf forest	5. Mixed forest	6. Closed shrublands	7. Open shrublands	8. Woody savannas	9. Savannas	10. Grasslands	11. Permanent wetlands	12. Croplands	14. Cropland/natural vegetation mosaic	15. Snow and ice	16. Barren or sparsely vegetated	
0.2	3423	6339	728	45	4902	12,520	381	12,551	5690	483	23,759	651	14,324	10,873	21	1256	97,946
0.4	6886	12,435	1436	91	9104	25,585	668	25,103	5690	724	47,518	1303	28,649	21,747	38	2435	189,412
0.6	10,270	19,018	2215	136	15,932	36,779	1336	39,448	5690	1449	69,693	1882	42,351	32,621	64	3896	282,780
0.8	13,613	24,870	2953	182	21,243	49,038	1336	55,227	5690	1449	95,036	2420	57,298	43,495	96	4870	378,816
1.0	17,217	32,331	3544	234	21,243	65,385	2672	69,034	5690	1449	116,155	3388	69,577	65,243	96	6493	479,751
1.5	25,451	46,188	5316	328	31,865	98,077	2672	92,045	5690	1449	174,233	4235	108,231	65,243	193	9740	670,956
2.0	34,434	64,663	7594	410	63,731	117,693	2672	138,068	5690	1449	261,349	5646	139,154	130,486	193	12,987	986,219

Table 4
Average percentage of study area (Fig. 1) land cover class training data samples (n_i^*) (Table 3) that are replaced by local training pool data over the 511 3 × 3 GWELD tile neighborhoods considering different percentages (p) of the study area 500 m pixels.

p (%)	Land cover class																Row mean
	0. water	1. Evergreen needleleaf forest	2. Evergreen broadleaf forest	3. Deciduous needleleaf forest	4. Deciduous broadleaf forest	5. Mixed forest	6. Closed shrublands	7. Open shrublands	8. Woody savannas	9. Savannas	10. Grasslands	11. Permanent wetlands	12. Croplands	14. Cropland/Natural Vegetation mosaic	15. Snow and ice	16. Barren or sparsely vegetated	
0.2	24.073	12.814	4.504	23.596	8.239	15.178	1.847	9.414	0.000	1.384	19.666	8.293	25.271	10.572	0.438	9.620	10.932
0.4	17.231	10.942	3.686	15.251	5.917	12.986	1.312	7.296	0.000	0.963	15.337	6.261	18.562	6.332	0.227	8.309	8.163
0.6	14.377	9.752	3.385	11.888	3.862	11.775	0.675	5.500	0.000	0.000	12.548	5.237	15.459	4.382	0.131	6.508	6.592
0.8	12.817	8.903	3.213	9.568	2.848	10.548	0.675	4.325	0.000	0.000	10.540	4.539	13.005	3.229	0.084	5.841	5.633
1.0	11.657	7.963	3.116	7.796	2.848	9.021	0.000	3.616	0.000	0.000	9.368	3.744	11.583	1.746	0.084	4.827	4.836
1.5	10.046	6.461	2.776	5.741	1.672	6.849	0.000	2.777	0.000	0.000	6.992	3.079	8.714	1.746	0.000	3.399	3.766
2.0	8.990	5.028	2.311	4.389	0.000	5.863	0.000	1.640	0.000	0.000	4.905	2.379	7.350	0.000	0.000	2.558	2.838

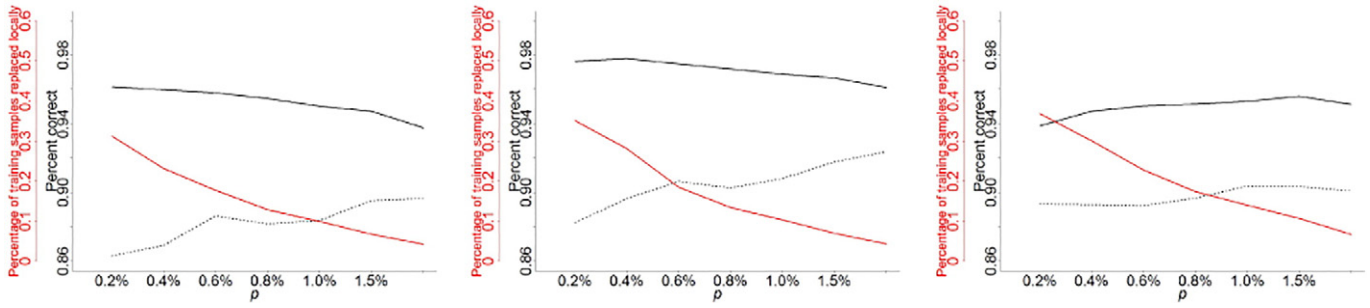

Fig. 3. Sensitivity of random forest classification accuracy to the percentage of the North American MODIS land pixels sampled (p) for three GWELD tiles. Overall classification accuracies derived from the tile OOB samples for the single random forest classification (black dotted lines) and the locally adaptive random forest classification (black lines). The red lines show the percentage of the single random forest training data replaced with local training pool samples used in the locally adaptive classification. The California hh08vv05.h2v2 tile (*left*) is covered predominantly by MCD12Q1 classes cropland (57.0%), open shrublands (12.4%) and grasslands (12.0%), the Oregon hh09vv04.h2v5 tile (*middle*) is mainly grasslands (96.9%), and the Minnesota hh11vv04.h4v1 tile (*right*) is covered by mixed forest (39.1%), cropland/natural vegetation mosaics (27.1%), and cropland (19.1%).

with p up to 1.5%, although there is relatively little variation in the single or locally adaptive random forest classification accuracy (<1.5% for the different illustrated p values). As both classification methods provide accuracies of about 90% this suggests that the local Minnesota class training data are not dissimilar to the class training pool data for the rest of the study area.

Importantly, for all the illustrated Fig. 3 examples, the locally adaptive classifications are systematically more accurate (by 3.75% to 9.84%) than the single random forest classifications. The optimal p value for the locally adaptive classification is unknown *a priori* as it is dependent on the local training data availability and quality relative to the training pool. A data driven solution to find the optimal p value is suggested. For example, one where the locally adaptive (Fig. 3, black solid line) classification results are generated for different p values and the value that maximizes the classification accuracy is selected. However, this is computationally prohibitive as it means classifying the data multiple times. Consequently, in this study a single fixed p value was used.

In this study a 1.0% p value was used for the single and locally adaptive random forest classifications. This provided 479,751 training pixels balanced among classes, as Eq. (3), according to the study area MCD12Q1 2010 class proportions (Table 3). Thus, 479,751 training pixels were selected from the training pool and used for the single random forest classification. The locally adaptive random forest classification also used these 479,751 training pixels but replaced them on a class basis with available local training pool data. For $p = 1.0\%$ three of the classes have no locally updated training samples updated over the study area 3 × 3 GWELD tile neighborhoods (Table 4). Specifically, woody savannas (class 8) never has locally updated training samples for any p and savannas (class 9) and closed shrublands (class 6) only have locally updated data when p is not >0.4% and 0.8% respectively. For the remainder of the classes, when $p = 1.0\%$, the average percentage of the study area land cover class training samples replaced by local training pool samples varies among the classes from 0.084% to 11.657% and when considering all classes the average is 4.836% (Table 4). This was judged to be a suitable compromise as using a lower p reduced the amount of training data used (Table 3). We note that for both the single and locally adaptive random forest classifications using 479,751 training pixels provides a considerable training data set. For comparison, only 91,433 training samples were used to generate a global nine class 30 m Landsat land cover classification (Gong et al., 2013; Yu et al., 2013).

4.2. Single random forest 30 m land cover classification

Fig. 4 shows the single random forest 30 m classification generated using the 500 trees applied to all 511 GWELD tiles. There were a total of 11,967,033,918 30 m pixels of which 11,951,131,523 (99.87%) were classified, while 15,902,395 (0.13%) were unclassified because there were less than five valid observations over the 36 months of GWELD data. Superimposed on the GWELD classification, shown in red, is the NLCD-based CONUS 30 m urban mask. At this synoptic scale, the classification appears geographically plausible and is similar to the 2010 MCD12Q1 500 m classification (Fig. 8, Section 4.4).

Fig. 5 shows the CONUS classification confidence image, i.e., the proportion of times over the 500 trees that each pixel was classified as the majority class (Section 3.4.2). The confidence values do not measure accuracy but rather provide an indication of the classification quality. Considering the non-missing pixels the confidence varies from 0.13 to 1.0 with a median of 0.77 (mean of 0.749) which indicates reasonable confidence for 16 classes. The spatial distribution of the classification confidence data is of interest. High confidences typically occur over water bodies which is expected, as water is spectrally quite different to other classes in Landsat data (Sheng et al., 2016), and also over the extensive deserts and drylands across the CONUS which is presumably because bright soils are markedly spectrally different and more temporally stable than the other classes. Low classification confidence is apparent over certain agricultural regions reflecting the diversity of crop types that can have complex spectral and temporal signatures (Chang et al., 2007; Johnson and Mueller, 2010; Yan and Roy, 2015). The urban mask regions (red in Fig. 4) are quite apparent in the classification confidence image and typically have the lowest confidence (Fig. 5), for example, over Chicago that is located on the south west corner of Lake Michigan (detail in Fig. 2). This is expected as the GWELD classification did not include an urban class. A total of 2.37% of the CONUS pixels were overwritten with the NLCD-based 30 m urban mask. It is interesting to note that these urban pixel locations were not frequently classified as barren or sparsely vegetated (1.77%, class 16) but rather were more frequently classified as croplands (30.16%, class 12), grasslands (23.67%, class 10), cropland/natural vegetation mosaic (20.59%, class 14), and open shrublands (13.09%, class 7). This likely reflects, except for desert cities and dense central business districts, the documented and nationally variable presence of vegetation within U.S. urban areas (Pataki et al., 2006).

Table 5 reports the single random forest classification confusion matrix and the class producer's and user's accuracies that quantify the level of agreement between the classification and the rigorously filtered MCD12Q1 class labels present in the bootstrapped training pool data. The overall classification level of agreement is high, with a 93.13% percent correct and 0.9195 kappa. The user's and producer's accuracies were both >90% for the following ten classes: water (class 0), evergreen needleleaf forest (class 1), mixed forest (class 5), closed shrublands (class 6), open shrublands (class 7), grasslands (class 10), permanent wetlands (class 11), croplands (class 12), snow and ice (class 15), and barren or sparsely vegetated (class 16). The user's and producer's accuracies were >83% for all the other classes except for deciduous needleleaf forest (class 3) and woody savannas (class 8). These results

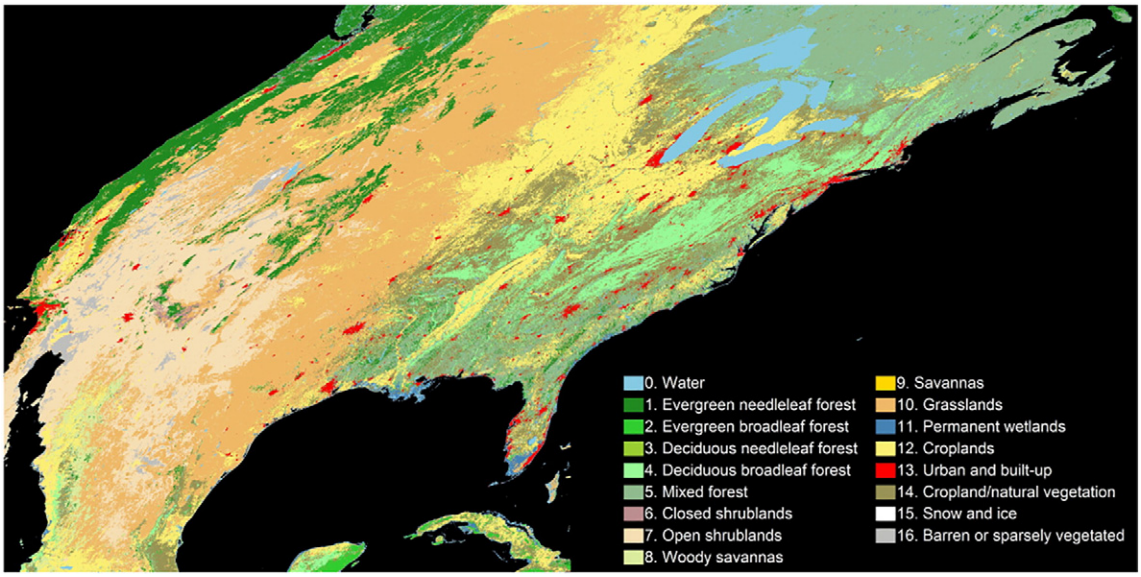

Fig. 4. Single random forest 30 m land cover classification derived using 479,751 training pixels drawn balanced to the study area MCD12Q1 class proportions from the training pool (p = 1.0%, Table 3). The urban and built-up class (red) over the CONUS was defined by the static NLCD-based CONUS 30 m urban mask (Section 2.4).

indicate quite reasonable classification accuracies. Deciduous needleleaf forest (class 3) had a markedly poor accuracy with a 33.3% user's and 0.4% producer's accuracy. Notably, this was the one class for which the training pool data were not generated using the spatial filter (Section 3.2) because no good quality consistent deciduous needleleaf forest MCD12Q1 500 m pixels remained after the filtering. We note also that this class was frequently (143 of 234 training samples) confused with the mixed forest (class 5) which has a similar class definition. Woody savannas (class 8) had producer's and user's accuracies of 72.0% and 90.3% respectively and was often classified as grasslands (class 10), croplands (class 12), or cropland/natural vegetation mosaic (class 14). Again this may be because of the relative scarcity of training data for this class (Table 3). The results in Table 5 are quite similar but higher than those reported for the global MODIS collection 5 MCD12Q1 product (Table 4 in Friedl et al., 2010).

4.3. Locally adaptive random forest 30 m land cover classification

Fig. 6 shows the locally adaptive random forest 30 m classification. The overall level of agreement between the classification and the rigorously filtered MCD12Q1 class labels present in the bootstrapped training pool data is about 2% higher than for the single random forest classification, with a 95.44% correct and 0.9443 kappa. At this synoptic scale there is no evident difference between these results and the Fig. 4 classification results. The classification confidence also appears similar to Fig. 5 and so is not illustrated. Considering all the non-missing pixels the confidence varies from 0.13 to 1.0 with a median of 0.79 (mean 0.761) which is on average slightly higher than for the single random forest classification.

Fig. 7 illustrates two example full resolution 1000 × 1000 30 m pixel subsets of the locally adaptive classification selected over regions of predominantly cropland (left) and deciduous broadleaf forest (right).

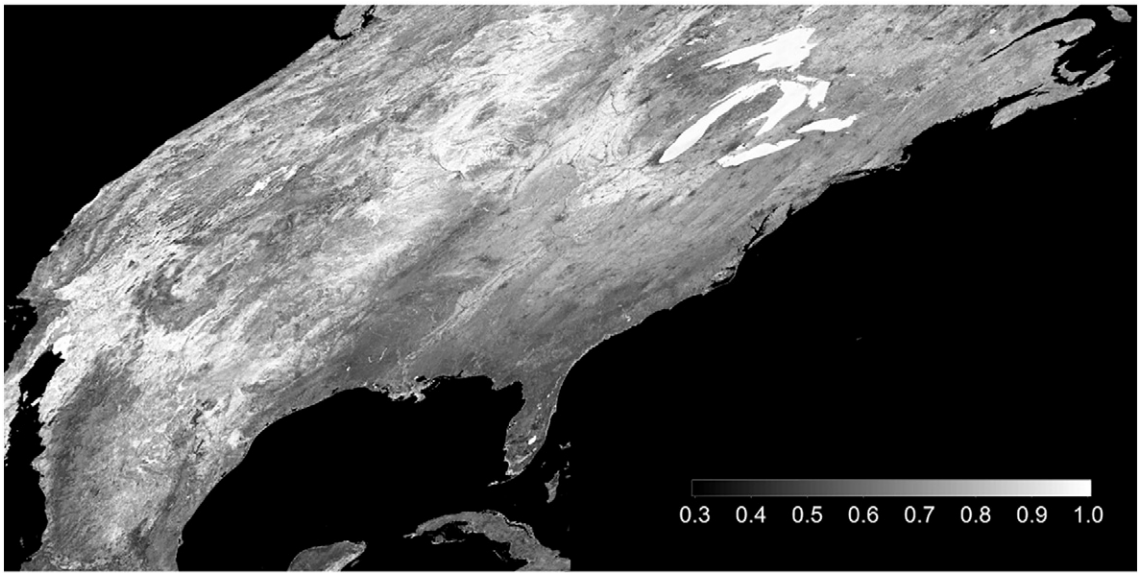

Fig. 5. Classification confidence map (highest confidence 1.0 shown in white) for the single random forest classification illustrated in Fig. 4.

Table 5
Single random forest 30 m classification (Fig. 4) confusion matrix results. A total of 479,751 classified "out-of-bag" (OOB) samples for 16 classes (Table 1) were used to derive the confusion matrix counts. The class producer's and user's accuracies quantify the level of agreement between the classification and the rigorously filtered MCD12Q1 class labels present in the bootstrapped training pool data.

		Training land cover class																
		0. Water	1. Evergreen needleleaf forest	2. Evergreen broadleaf forest	3. Deciduous needleleaf forest	4. Deciduous broadleaf forest	5. Mixed forest	6. Closed shrublands	7. Open shrublands	8. Woody savannas	9. Savannas	10. Grasslands	11. Permanent wetlands	12. Croplands	14. Cropland/natural vegetation mosaic	15. Snow and ice	16. Barren or sparsely vegetated	Row sum	User's accuracy (%)
Classified land cover class	0	17,193	9	0	3	0	2	0	0	0	0	5	16	0	0	0	0	17,228	99.8
	1	4	30,506	69	26	2	1047	63	11	195	22	71	24	11	16	0	0	32,067	95.1
	2	0	8	3059	0	19	84	0	0	33	0	0	6	4	17	0	0	3230	94.7
	3	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	3	33.3
	4	0	3	73	1	17,807	1165	0	0	47	0	0	0	2	1944	0	0	21,042	84.6
	5	4	1403	263	143	1370	61,475	0	0	70	0	4	89	20	1470	0	0	66,311	92.7
	6	0	81	0	0	0	0	2493	33	32	4	19	0	4	1	0	0	2667	93.5
	7	1	7	0	0	0	0	21	66,381	15	7	2250	0	199	0	1	107	68,989	96.2
	8	0	141	1	10	5	50	22	15	4096	42	72	4	26	51	0	0	4535	90.3
	9	0	12	0	0	1	0	1	13	34	1270	9	0	1	0	0	0	1341	94.7
	10	2	68	0	4	2	2	37	2414	382	103	111,933	9	1490	90	2	8	116,546	96.0
	11	13	29	2	26	2	58	1	0	5	1	95	3232	48	42	0	0	3554	90.9
	12	0	18	3	3	42	15	18	89	407	0	1512	2	63,479	4236	0	0	69,824	90.9
	14	0	45	74	17	1993	1487	16	0	374	0	156	6	4293	57,375	0	0	65,836	87.1
	15	0	0	0	0	0	0	0	0	0	0	1	0	0	0	92	1	94	97.9
	16	0	0	0	0	0	0	0	78	0	0	28	0	0	0	1	6377	6484	98.3
Column sum		17,217	32,331	3544	234	21,243	65,385	2672	69,034	5690	1449	116,155	3388	69,577	65,243	96	6493	479,751	
Producer's accuracy (%)		99.9	94.4	86.3	0.4	83.8	94.0	93.3	96.2	72.0	87.6	96.4	95.4	91.2	87.9	95.8	98.2		

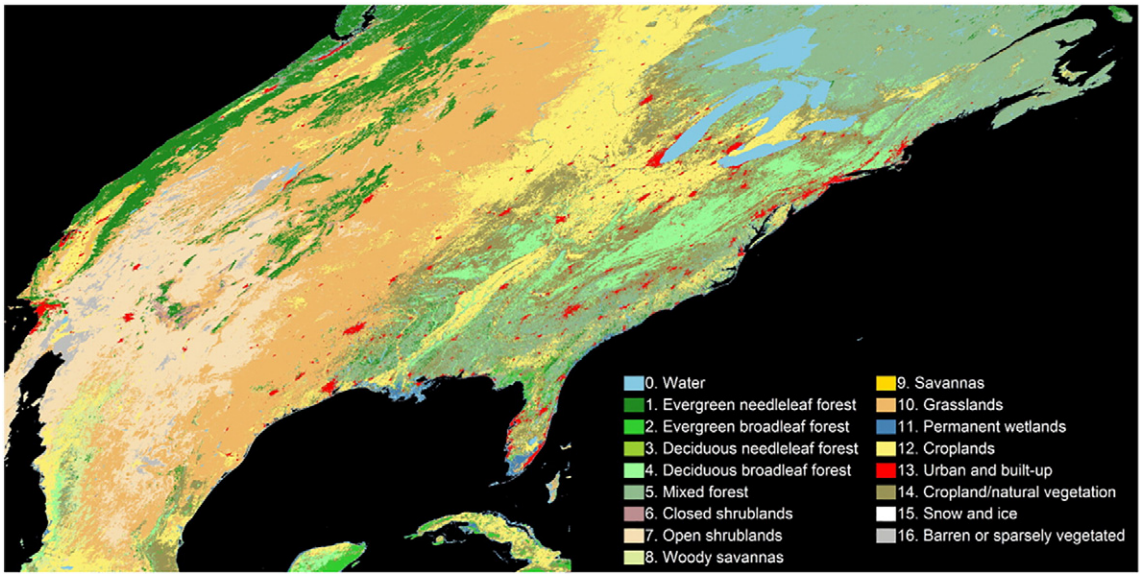

Fig. 6. Locally adaptive random forest 30 m land cover classification derived using 479,751 training pixels with locally replaced training pixels from each 3 × 3 GWELD tile area ($p = 1.0\%$, Table 4). The urban and built-up class (red) was defined by the static NLCD-based CONUS 30 m urban mask (Section 2.4).

These examples were selected specifically because they are centered on the intersection of four neighboring GWELD tiles (the corner of each tile occupies a subset quadrant) and were classified using largely different amounts of locally updated training data. There is no tile classification boundary apparent in these two examples. This is because 3 × 3 GWELD tiles were used for the local training sample updating and the resulting random forest was used to classify the central tile.

Table 6 reports the CONUS locally adaptive classification confusion matrix and the producer's and user's accuracies that quantify the level of agreement between the classification and the rigorously filtered MCD12Q1 class labels present in the bootstrapped training pool data. As for the single random forest results (Table 5) the user's and producer's accuracies were both >90% for ten of the sixteen classes. Considering both the user's and producer's accuracy the croplands class (12) has the greatest accuracy improvement in the locally adaptive classification compared to the single random forest classification with 2.7% and 5.6% higher producer's and user's accuracy. This is likely because croplands are spatially and temporally highly variable. For example, the USDA 30 m Cropland Data Layer (CDL) product maps >100 CONUS crop types (Boryan et al., 2011; Johnson and Mueller, 2010) and multiple crops can be rotated over several years (Plourde et al., 2013). Moreover, the same crop type in different regions may have different soils, growing seasons and vigor, and be subject to different crop management practices and weather that effect its discrimination using satellite data (Doraiswamy et al., 2004; Boryan et al., 2011; Yan and Roy, 2015).

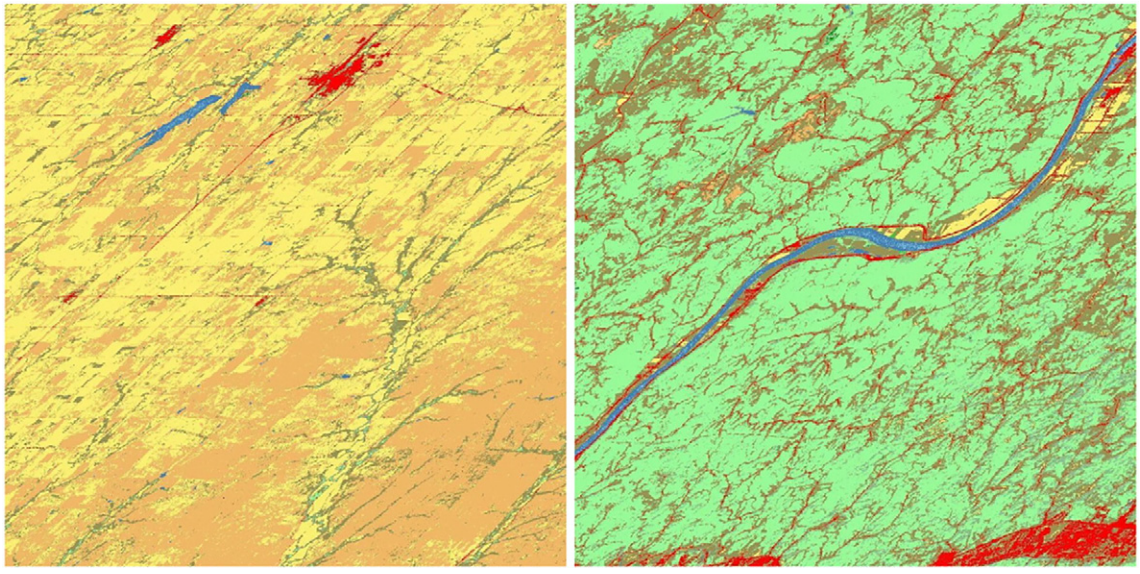

Fig. 7. Locally adaptive random forest 1000 × 1000 30 m land cover classification results for Herington, Kansas (left) and for the Crown City (Ohio) and Mill Creek (West Virginia) wildlife management areas to the north of Milton and Hurricane (West Virginia) (right), each centered on the intersection of four adjacent GWELD tiles. The four tiles (left) hh10vv05.h2v0, hh10vv05.h3v0, hh10vv05.h2v1, hh10vv05.h3v1 were classified using 21.75%, 2.47%, 13.77% and 0.98% of the training data samples replaced locally from the training pool. The four tiles (right) hh11vv05.h3v0, hh11vv05.h4v0, hh11vv05.h3v1 and hh11vv05.h4v1 were classified using 14.70%, 2.81%, 15.74% and 2.11% of the training data samples replaced locally from the training pool. The class colors and labels are provided in Fig. 6 and Table 1 respectively.

Table 6
Locally adaptive random forest 30 m classification (Fig. 6) confusion matrix results. A total of 2,222,105 classified "out-of-bag" (OOB) samples for 16 classes (Table 1) were used to derive the confusion matrix counts considering the training samples located within each of the 511 GWELD tiles. The class producer's and user's accuracies quantify the level of agreement between the classification and the rigorously filtered MCD12Q1 class labels present in the bootstrapped training pool data.

		Training land cover class																
		0. Water	1. Evergreen needleleaf forest	2. Evergreen broadleaf forest	3. Deciduous needleleaf forest	4. Deciduous broadleaf forest	5. Mixed forest	6. Closed shrublands	7. Open shrublands	8. Woody savannas	9. Savannas	10. Grasslands	11. Permanent wetlands	12. Croplands	14. Cropland/natural vegetation mosaic	15. Snow and ice	16. Barren or sparsely vegetated	Row sum	User's accuracy (%)
Classified land cover class	0	107,891	53	0	26	0	12	0	0	0	0	23	98	0	0	1	0	108,104	99.8
	1	34	164,321	288	165	2	3940	68	20	183	17	340	37	53	30	0	0	169,498	96.9
	2	0	7	11,236	0	45	173	0	0	18	0	0	5	2	23	0	0	11,509	97.6
	3	7	5	0	43	0	17	0	0	0	0	0	1	0	0	0	0	73	58.9
	4	0	4	147	1	43,757	4439	0	0	59	0	0	0	15	3230	0	0	51,652	84.7
	5	64	5628	899	807	3190	364,659	0	0	82	0	17	231	106	2796	0	0	378,479	96.3
	6	0	202	0	0	0	0	2484	70	32	4	90	0	16	0	0	0	2898	85.7
	7	0	21	0	0	0	0	29	186,799	16	6	7588	0	1079	0	0	158	195,696	95.5
	8	0	520	2	38	9	243	21	21	4079	42	158	6	89	94	0	0	5322	76.6
	9	0	47	0	0	2	0	1	28	35	1282	38	0	3	0	0	0	1436	89.3
	10	10	250	0	38	6	19	36	4913	371	96	644,611	26	7979	200	5	21	658,581	97.9
	11	129	150	8	160	8	284	0	0	4	1	253	10,911	209	94	0	0	12,211	89.4
	12	0	84	1	40	71	101	14	263	429	1	5044	5	444,852	9775	0	0	460,680	96.6
	14	0	135	189	125	3934	7041	19	0	382	0	621	13	19,303	114,244	0	0	146,006	78.2
	15	1	0	0	0	0	0	0	0	0	0	6	0	0	0	93	1	101	92.1
	16	0	0	0	1	0	0	0	185	0	0	53	0	3	0	0	19,617	19,859	98.8
Column sum		108,136	171,427	12,770	1444	51,024	380,928	2672	192,299	5690	1449	658,842	11,333	473,709	130,486	99	19,797	2,222,105	
Producer's accuracy (%)		99.8	95.9	88.0	3.0	85.8	95.7	93.0	97.1	71.7	88.5	97.8	96.3	93.9	87.6	93.9	99.1		

The producer's accuracy is used to compare classification algorithms as it quantifies for a given class in the training data how many are classified correctly (Story and Congalton, 1986). The class producer's accuracies are comparable (within 1%) or greater for the locally adaptive random forest (Table 6 bottom row) than for the single random forest classification (Table 5 bottom row). Only the snow and ice (class 15) has lower (by >1%) producer's accuracy in the locally adaptive classification. The reason for this cannot be reliably interpreted however given the relatively small number of snow and ice samples considered. The user's accuracy quantifies how often a class in the classification is present on the surface (assuming that the training data capture the surface class distributions) (Story and Congalton, 1986). Direct comparison of the class user's accuracy for the two classification methods cannot be undertaken reliably due to the different number and spatial sampling of "out-of-bag" (OOB) samples considered between the methods. Of note however, is that six of the sixteen classes have greater (by >1%) user's accuracy in the locally adaptive classification (Table 6 right column) than the single random forest classification (Table 5 right column). The classes with lower locally adaptive user's accuracy typically have a low average percentage of training samples replaced by the local training pool data (Table 4). In particular, no local training data were available for closed shrublands (class 6), woody savannas (class 8) and savannas (class 9) and these classes have the smallest user's accuracies (85.7%, 76.5% and 89.3% respectively) for the locally adaptive classification compared to the single random forest classification (93.5%, 90.3% and 94.7% respectively).

4.4. Comparison of the 30 m Landsat and 500 m MODIS land cover classifications

Fig. 8 illustrates the 2010 MCD12Q1 500 m land cover product which at this scale is generally consistent with the single (Fig. 4) and locally adaptive (Fig. 6) random forest 30 m classifications. Some regional differences are apparent however. In particular, North Dakota and the Canadian border region were classified as cropland (class 12) by MCD12Q1 (Fig. 8) but also as grassland (class 10) in the 30 m classifications (Figs. 4 and 6). This northern great plain region is characterized by a mixture of grassland and croplands (Sleeter et al., 2013). Another apparent synoptic scale difference is in the northwest Canadian boreal forest that was classified by MCD12Q1 as evergreen needle leaf forests (class 1) and mixed forests (class 5) but classified predominantly as mixed forests in the 30 m classifications (class 5). This region is dominated by needle leaf forests but also has extensive patches of herbaceous and grass species in forest disturbance (logging and fire) areas (Hermosilla et al., 2016; Bartels et al., 2016).

Figs. 9 and 10 illustrate two example 27 × 27 km subsets with six panels showing the MODIS MCD12Q1 500 m classification (top left), the 30 m single random forest classification (middle left), and the 30 m locally adaptive classification (bottom left) and the corresponding 30 m classification confidence maps (middle right and bottom right). For geographic context a true color 30 m GWELD 2010 annual composite is also shown (top right). These examples are in the CONUS and so include the NLCD urban mask (red classified pixels). The spatial resolution differences between the MODIS 500 m and the Landsat 30 m classifications are striking. Areas that are classified into a small number of 500 m MCD12Q1 land cover classes contain other additional classes that are captured at 30 m. This illustrates why using only local training data defined from MCD12Q1 is inappropriate for 30 m classification as local 500 m training data may not capture classes that are actually present.

Fig. 9 shows a region of S.E. Oregon that is classified by MCD12Q1 into six classes, predominantly grasslands (class 10, 94.8%) and open shrublands (class 7, 4.3%). In the 30 m classifications these class proportions are quite different with notably less grassland (82.3% and 61.2% in the locally adaptive and single random forest classifications respectively) and more open shrublands (12.6% and 33.4% in the locally adaptive and single random forest classifications respectively) and a minority of 30 m pixels classified into five new classes. The green circular pivot irrigation field (that appears elliptical due to the sinusoidal projection) in the north west corner of the true color GWELD 2010 annual composite is correctly classified as cropland (class 12) in both the 30 m classifications but is under-represented in MCD12Q1 likely due to the coarser 500 m resolution. Similarly, the diagonal road apparent in the S.E. corner of the 30 m classifications (red) is not present in MOD12Q1. Across the subset the locally adaptive 30 m classification appears more coherent than the single random forest classification, particularly because the open shrubland (class 7) classified pixels are spatially more coherent and replace pixels classified in the single random forest classification as grassland (class 10). Shrublands and grasslands that coexist are often difficult to classify reliably as they can be spectrally similar (Wickham et al., 2013; Gibbes et al., 2010). Notably, the classification confidence for the shrublands and grasslands classes is higher for the locally adaptive classification than the single random forest classification. In both classification confidence data sets, and also in the single random

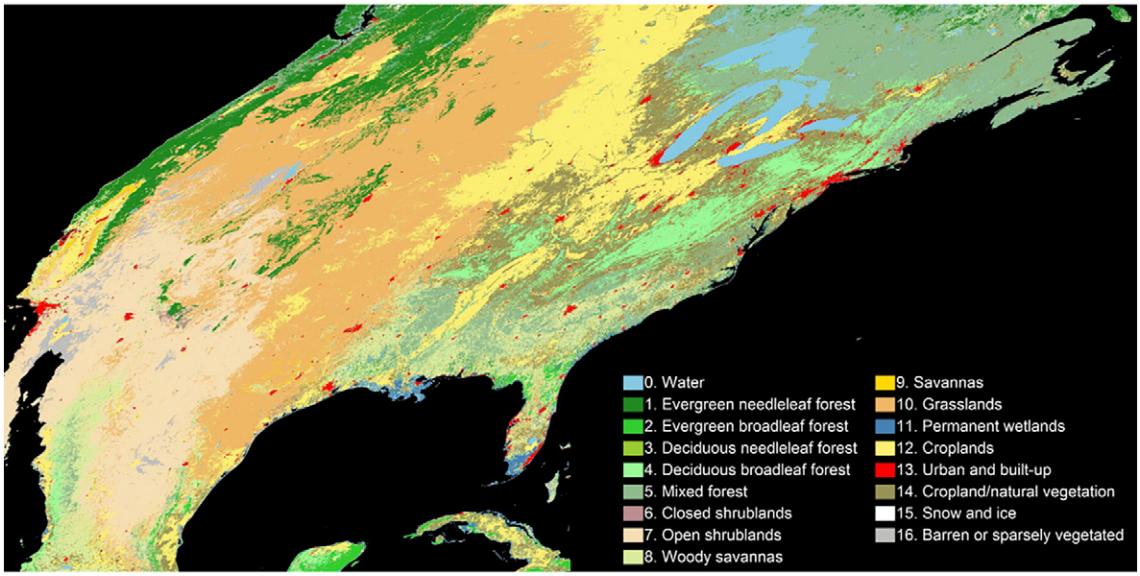

Fig. 8. Collection 5 2010 MCD12Q1 500 m land cover classification (Friedl et al., 2010).

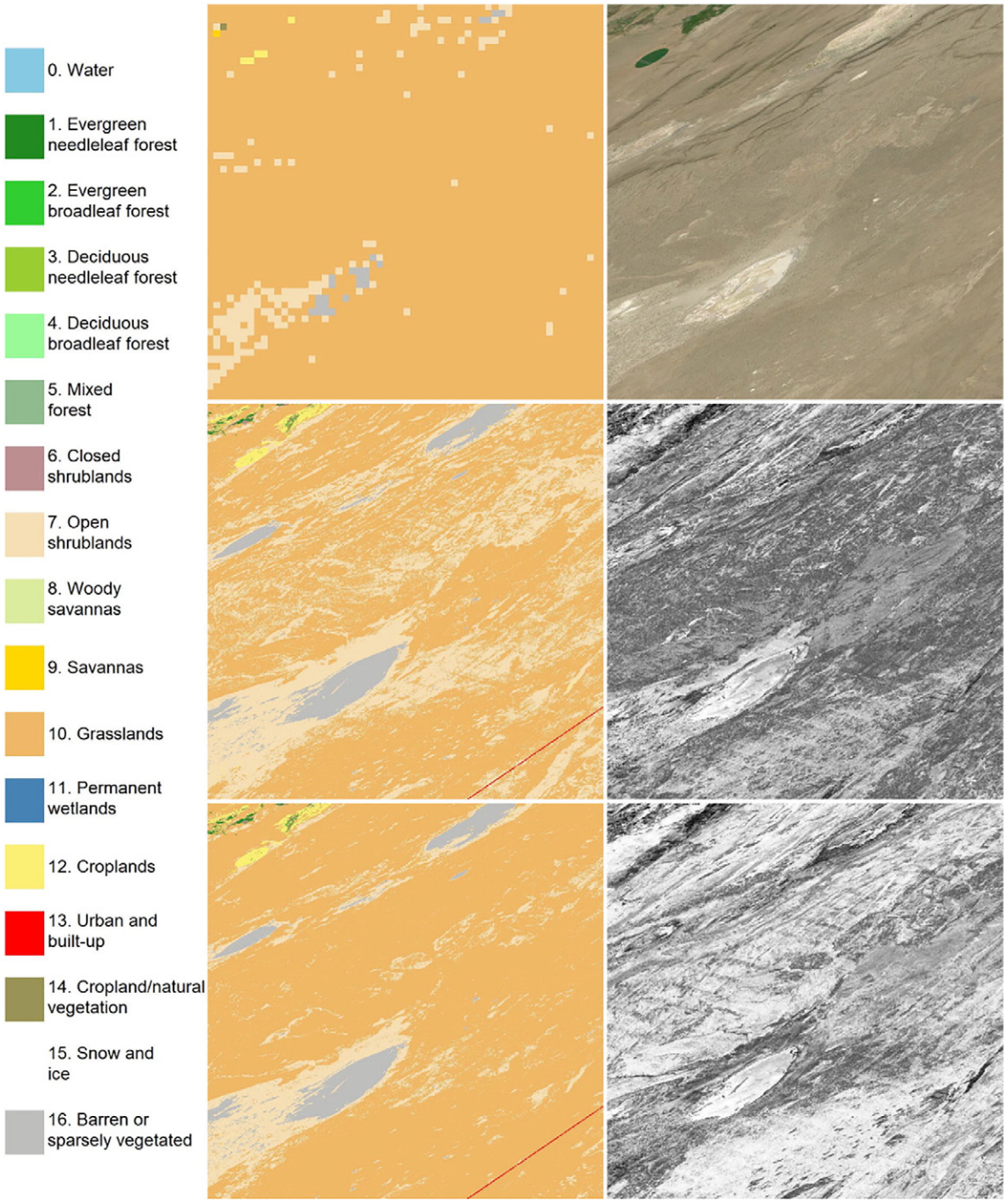

Fig. 9. Detailed comparison of MCD12Q1 500 m 2010 classification (top left), the single random forest classification (middle left), the locally adaptive classification (bottom left) and the corresponding 30 m classification confidence maps (middle right and bottom right). For geographic context a true color GWELD 2010 annual composite is also shown (top right). Example for a 27 × 27 km (900 × 900 30 m, 54 × 54 500 m) pixel area centered on 42.6278°N 118.1629°W near Andrews, Oregon (falling in GWELD tile hh09vv04.h2v5).

forest classification, a faint striping pattern aligned approximately N.W. to S.E. is evident. This is due to the Landsat-7 ETM + scan line corrector failure that introduced across-track stripes of missing pixels with different spatial phases through time (Markham et al., 2004). Although the 30 m classifications were generated from temporal Landsat metrics that reduce the impact of this issue, stripes become apparent where there are more discarded cloudy pixels and fewer available observations in the time series (Lindquist et al., 2008). For each subset the mean classification confidence is 0.8136 and 0.6993 for the locally adaptive and single random forest classification respectively, indicating that the locally adaptive classification is more reliable.

Fig. 10 shows similar results as Fig. 9 but for a region in central California. The subset is classified by MCD12Q1 into four classes: cropland (class 12, 96.8%), urban and built-up (class 13, 2.4%), grasslands (class 10, 0.5%) and open shrublands (class 7, 0.3%). In the 30 m classifications these class proportions are quite different with many more classes reflecting the heterogeneous landscape that is apparent in true color GWELD 2010 annual composite. The evergreen forest classes and much of the road network (red) are absent in MCD12Q1, likely due to the coarser 500 m resolution. The urban areas and roads are by definition the same for both 30 m classifications as they are defined independently by the NLCD derived urban mask. Small ephemeral water bodies,

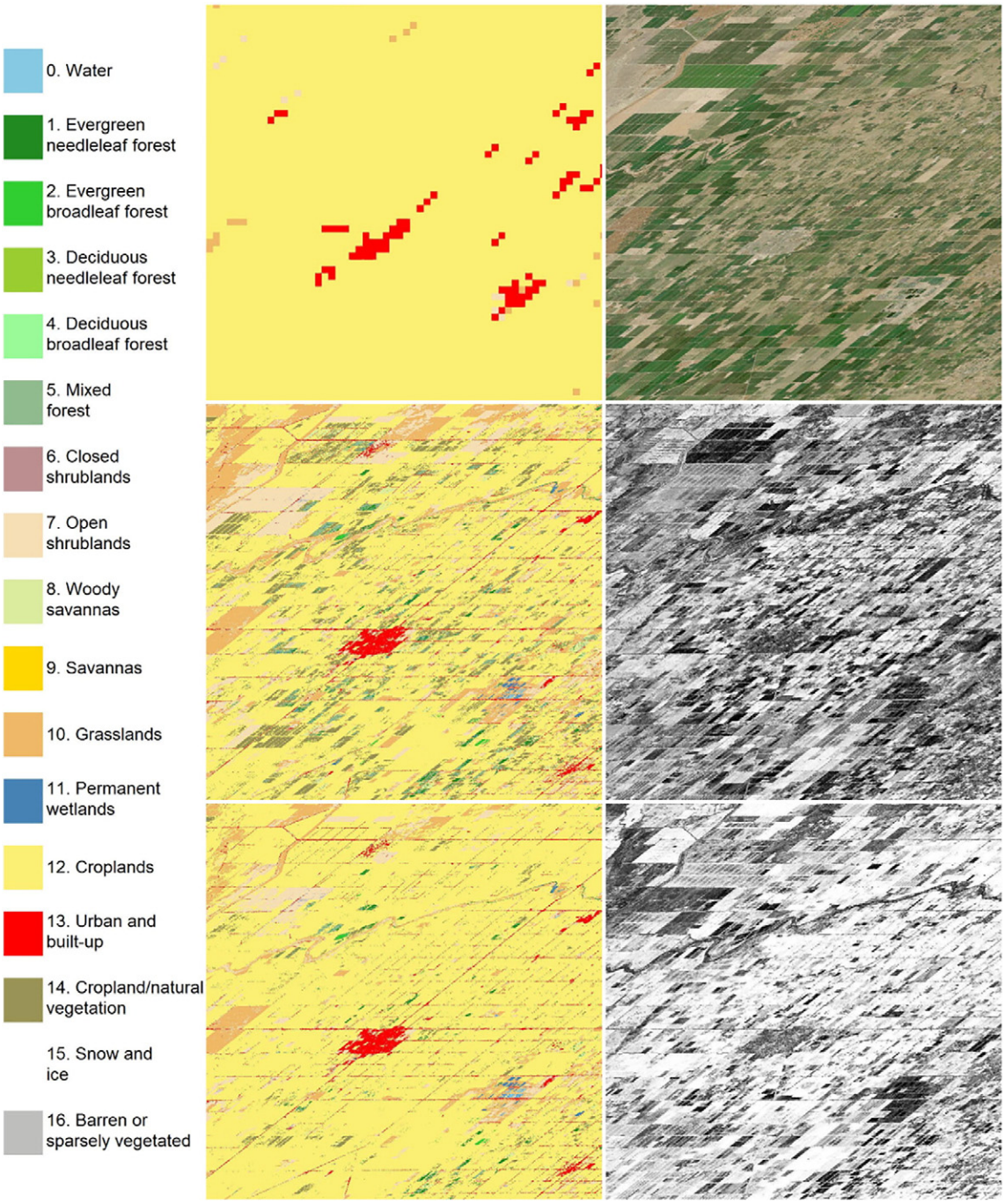

Fig. 10. As Fig. 9 for a 27 × 27 km area centered on 36.7517°N 120.0795°W near Kerman, California (falling in GWELD tile hh08vv05.h2v2).

such as near Dry Creek Canal in the S.E. part of the subset, are classified as water (class 0, blue) in the 30 m classifications but are misclassified as urban in MCD12Q1. In general, the differences between the 30 m land cover classifications is because the majority of the crop/natural vegetation mosaic (class 14) and also some of the evergreen forests (classes 1 and 2) single random forest classification pixels are classified as croplands (class 12) in the locally adaptive classification. Inspection of the true color GWELD 2010 annual composite indicates that the majority of the landscape is agriculture. As noted earlier, croplands are spatially and temporally highly variable and the use of local training data captures this variability better than the single random forest classification derived for all the study area. As in Fig. 9, the locally adaptive classification has higher classification confidence (mean 0.8386) than for the single CONUS random forest classification (mean 0.7118).

Fig. 11 illustrates for each of the 17 land cover classes the CONUS proportion of the resampled 30 m MCD12Q1 pixels classified as another (or the same) class in the 30 m locally adaptive random forest classification. The proportions are derived as Eq. (4). The 500 m MCD12Q1 pixels may be classified as several other classes in the 30 m locally adaptive random forest classifications due to misclassification inaccuracies in one or both land cover products and due to the different spatial arrangement of the land cover at 30 m and 500 m. The classes with the most mixed proportions, evergreen broadleaf forest (class 2), deciduous needleleaf forest (class 3), woody savannas (class 8) and savannas (class 9), are all poorly classified with low (<90%) producer's or user's accuracy (Section 4.3). Across the CONUS for 13 of the 17 classes the majority 30 m land cover class falling within each 500 m pixel is the same as MCD12Q1 class. The majority class proportions for these classes

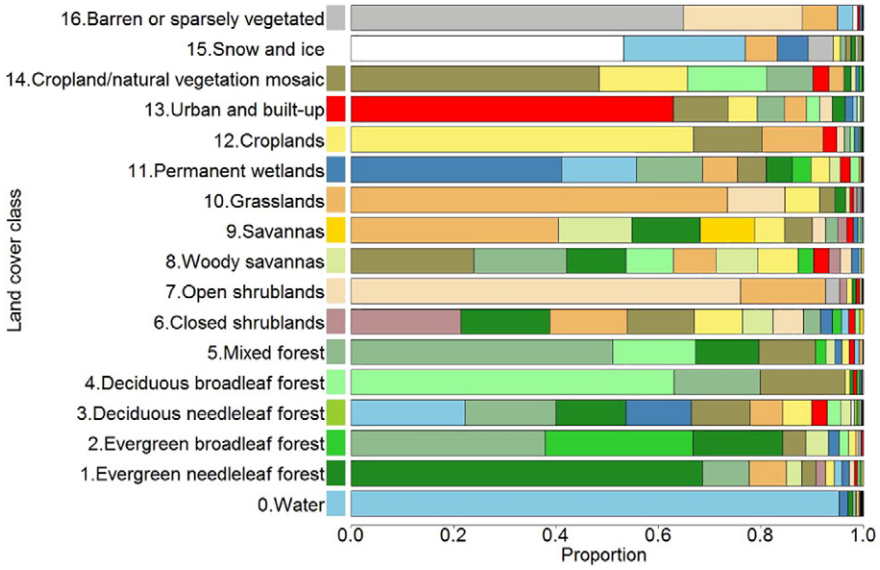

Fig. 11. For each of the 17 land cover classes (rows) the CONUS proportions of the resampled 30 m MCD12Q1 pixels classified as another (or the same) class in the 30 m locally adaptive random forest classification are shown by the colored horizontal area. The proportions are derived as Eq. (4).

range from a maximum of 95.2% (class 0, water) to 21.4% (class 6, closed shrubland). The low closed shrubland majority proportion is expected as this class is easily confused with grasslands and also with forest as it is a transitional class between grassland and forest. Similarly, the wetland (class 11) and cropland/natural vegetation mosaic (class 14) classes have the next lowest majority class proportions (41.14% and 48.47%) and these classes are known to be difficult to classify reliably (Ozesmi and Bauer, 2002; Herold et al., 2008). The majority (62.9%) of the CONUS MCD12Q1 urban and built-up pixels were classified as developed high, medium or low urban density in the 30 m NLCD product (Fig. 11). The purpose of this study is not to compare the MCD12Q1 and NLCD urban classification results, however, the urban class is included in Fig. 11 to reduce bias because it accounts for a greater CONUS area than several of the other land cover classes (1.52% and 2.37% of the CONUS were classified as urban by the MCD12Q1 and the NLCD urban mask respectively).

5. Discussion

Global land cover classification is challenging due to the large volume data pre-processing required and the cost and difficulty of collecting representative training data that enable classification models to be globally consistent and locally reliable (Loveland et al., 2000; Bartholomé and Belward, 2005; Friedl et al., 2010; Gong et al., 2013). With the availability of analysis ready data, such as provided by the global Web-Enabled Landsat data (GWELD) products, and the development of non-parametric classifiers, the major challenge is training data collection. Over large areas the optimal training data sampling needed to provide a given classification accuracy is nearly always unknown. Consequently, land cover mapping is usually undertaken in an iterative manner by repeated training data collection and refinement, satellite data classification, and then classification accuracy assessment, until an acceptable classification accuracy is obtained (Egorov et al., 2015). In this study a novel methodology to classify large volume Landsat data using high quality training data derived from the MODIS land cover product was demonstrated. This is advantageous as it (i) enables the classification to be undertaken in an automated manner without the need for interactive and manual training data collection and refinement, (ii) provides a large geographically distributed training data set, and (iii) results in the generation of a 30 m Landsat land cover product with the same classification legend as the MODIS land cover product.

A training data pool was extracted from the MODIS land cover product by judicious quality and spatial filtering. Given the considerable training pool size a sample corresponding to 1% of the number of North America MODIS 500 m land cover pixels was selected. This provided 479,751 training samples, which we note is more than five times larger than the training sample set size used to make a recent global Landsat land cover classification (Gong et al., 2013; Yu et al., 2013). In addition, the training data selection was undertaken in a geographically systematic manner while ensuring that the selected class proportions were the same as the North America MODIS land cover product class proportions. This class proportion balancing is similar to the MODIS Collection 5 land cover product generation approach that utilized a Bayes Rule method to address training sample selection bias (Friedl et al., 2010) and to that used for updating the NLCD land cover product (Xian et al., 2009).

The detailed results illustrated in Figs. 9 and 10 are representative of the results across the study area in that the locally adaptive random forest classification was more coherent and had higher classification confidence than the single random forest classification. This and the higher overall and generally higher producer's and user's accuracies (that quantify the level of agreement between the classification and the filtered MCD12Q1 class labels present in the bootstrapped training pool data) underscore the utility of the locally adaptive random forest classification approach. Locally defined training data are inherently more likely to capture local spectral heterogeneity within and among classes. Clearly, however, the utility of the locally adaptive random forest classification depends upon the availability of local training data. In this study this was not an issue as the training pool data derived from MCD12Q1 were geographically well distributed and only 20 of the 511 North America GWELD tiles had no training pool data (of which 16 were coastal tiles with small land portions) and when considering 3×3 neighboring GWELD tiles all of the tiles except one had training pool data.

The accuracy of both the single and locally adaptive random forest land cover classifications was assessed by bootstrapping the random forest implementation to derive confusion matrix based metrics. The metrics quantified the level of *agreement* between the 30 m random classifications and the rigorously filtered MCD12Q1 class labels present in selected training pool data. This is not the same as an absolute classification accuracy assessment. However, we note that the filtering to select the MCD12Q1 class labels of the training pool data was quite rigorous (only 500 m pixels classified consistently as the same MODIS land cover class over the three years with classification confidence

>50% and quality assessment set as "good quality" were selected). Consequently, the selected training pixels were likely classified in the MODIS product more accurately than reflected in the overall 75% classification MCD12Q1 accuracy reported by Friedl et al. (2010).

The level of agreement for the locally adaptive random forest classification was higher (95.44%, 0.9443 kappa) than for the single random forest classification (93.13%, 0.9195 kappa). The class producer's and user's accuracies that quantify the level of agreement between the classification and the MCD12Q1 class labels present in the bootstrapped training pool data were also high. For both classifications no more than four of the 16 classes had either producer's or user's classification accuracies smaller than 85%. These high levels of agreement were not driven by the use of an excessive amount of training data as 1% of the North America 500 m and <0.005% of the 30 m pixels were used as training data. We also note that although random forest bootstrapping accuracy assessment is unbiased (Breiman, 2001) studies have observed that it may provide inflated accuracies due to the presence of noise and spatial auto-correlation in the training data (Kennedy et al., 2015; Millard and Richardson, 2015). In this study, noise (e.g., due to cloud shadow and residual atmospheric contamination, cloud, and BRDF effects) was expected to be reduced because of the use of 20%, 50% and 80% percentile temporal metrics and because the metrics were extracted from Landsat surface NBAR and normalized NBAR band ratios. Spatial auto-correlation was expected to be mitigated because the training data were sampled in a geographically systematic manner across North America and were selected nominally at least one 500 m pixel apart, which is a comparable distance to the 600 m training and testing sample block distance used in the NLCD urban layer accuracy assessment to avoid spatial autocorrelation bias (Yang et al., 2003).

Further research to validate the land cover classification results using independent samples is recommended (Olofsson et al., 2014). In particular, a systematic comparison of the land cover classification results with ground-based observations or higher spatial resolution interpreted satellite data is recommended, for example, as undertaken as part of the NLCD production process (Wickham et al., 2017). The developed GWELD 30 m land cover product is complementary to the NLCD product. However, the NLCD product uses the Anderson Level II classification system (Homer et al., 2004; Stehman et al., 2008) which is not the same as the MODIS IGBP classification system (Friedl et al., 2010) precluding their direct quantitative comparison.

Researchers are increasingly using all the available Landsat data rather than just select cloud-free images for applications and science (Roy et al., 2014a). An increasing availability of analysis ready Landsat data means that large area land cover classification will become more straightforward. For some users, especially those with limited internet data access and computer resources (Roy et al., 2010b), the scope of their land cover processing may be constrained. However, analysis ready data such as the GWELD and MODIS products are available in the high performance computing NASA Earth Exchange (https://nex.nasa.gov/nex/) environment, and Landsat data archives are being hosted and made available in commercial cloud processing environments.

The computational requirements to generate the North America classifications reported in this study were relatively modest. The GWELD metric generation and classification were implemented using custom C software. The processing was undertaken on a 64-bit Linux computer with 512 GB of memory and 32 cores and parallelized by tile with the processes distributed across the cores using the Open Multi-Processing application programming interface (http://www.openmp.org/). The total processing time was 91 h for the single random forest classification and 242 h for the locally adaptive random forest classifications respectively. Consequently, a global implementation is quite feasible as the methodology is automated. However, further research is merited, for example, concerning the outstanding need for a reliable 30 m global urban land cover product or classification methodology (Klotz et al., 2016), to consider the impact of scarce cloud-free Landsat data that occur in many parts of the world and periods of the Landsat record (Kovalskyy and Roy, 2013; Wulder et al., 2016), and the potential for more sophisticated Landsat time series based decomposition than empirical temporal metrics (Yan and Roy, 2015).

6. Summary

A novel methodology to classify large volume Landsat data using high quality training data derived from the MODIS land cover product was demonstrated for North America between 20°N and 50°N. Three years of MODIS 500 m land cover product and 39 temporal metrics extracted from three years of 30 m GWELD data derived from 55,978 Landsat 5 TM and Landsat 7 ETM+ images were used. The GWELD products provide consistently pre-processed surface nadir BRDF adjusted reflectance (NBAR) and are publically available so that other researchers may use them. They are stored in 30 m tiles that are nested within the standard 10° × 10° MODIS land product tiles so it is straightforward to compare the MODIS and GWELD products and undertake large volume tile-based processing. A data training pool was extracted from the MODIS land cover product by judicious quality and spatial filtering. All 17 MODIS IGBP land cover classes, except the Urban and built-up class, were considered. Each of the 500 m class labels was associated to only one 30 m Landsat pixel selected using a previously published method (Roy et al., 2016a) based on identifying the 30 m pixel that most closely matched the "metric centroid" of all the Landsat metric pixel values falling within the 500 m pixel. A total of 11,951,131,523 Landsat 30 m pixels across North America were classified. Two 30 m random forest classifications (run with default settings and using 500 trees) were generated. First, the 479,751 training samples were used to generate a single random forest classification that was applied to all the GWELD tiles. Second, for each GWELD tile, a locally adaptive random forest model was built and applied to the tile using the single random forest training data updated with other available local training pool samples. To ensure locally adaptive classification consistency across the tile boundaries the local training data were sampled from the training pool for 3 × 3 adjacent tiles to build a random forest that was then used to classify the central tile.

The North America 30 m land cover classification results appeared geographically plausible and at synoptic scale were similar to the 500 m MODIS land cover product. Classification confidence maps, defining the proportion of times over the 500 trees that each 30 m pixel was classified as the majority class, revealed high confidences over water bodies and within extensive deserts and drylands, reflecting the relative consistency of these classes. Low confidences were observed particularly over agricultural areas reflecting the diversity of North America crop types land covers. Detailed visual inspection revealed that the locally adaptive random forest classifications and associated classification confidences were generally more coherent than the single random forest classification results. The North America mean classification confidence was higher for the locally adaptive random forest classification (0.761) than for the single random forest classifications (0.749). The level of agreement between the 30 m classifications and the MODIS land cover product derived training data was assessed by bootstrapping the random forest implementation. The locally adaptive random forest classification had higher overall agreement (95.44%, 0.9443 kappa) than the single random forest classification (93.13%, 0.9195 kappa). These results indicate that a locally adaptive random forest classification approach should be used in preference to a single random forest classification.

The MODIS 500 m land cover product is available globally on an annual basis (Friedl et al., 2010) and so provides a very large source of land cover training data. It has been reprocessed three times and a new version (Collection 6) will be reprocessed to reflect improved sensor knowledge and improved input MODIS data and processing algorithms (Justice et al., 2002). The methodology described in this paper could be applied to other global land cover products such as the 1 km GLC2000 (Bartholomé and Belward, 2005) or the 300 m GlobCover product

(Bontemps et al., 2011). Similarly, the methodology could be applied to other Landsat-like moderate spatial resolution data, such as provided by Sentinel-2 (Drusch et al., 2012) and work is currently underway to process Sentinel-2 data into registration with Landsat data in GWELD tiles (Yan et al., 2016; Roy et al., 2016b) to enable their utility for multi-source land cover classification.

Acknowledgements

This research was funded by the NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) program (cooperative agreement NNX13AJ24A), by the U.S. Geological Survey Landsat science team (grant G12PC00069), and by the NASA Land Cover/Land Use Change Multi-Source Land Imaging Science Program (grant NNX15AK94G).

References

Bartels, S.F., Chen, H.Y., Wulder, M.A., White, J.C., 2016. Trends in post-disturbance recovery rates of Canada's forests following wildfire and harvest. For. Ecol. Manag. 361, 194–207.

Bartholomé, E., Belward, A., 2005. GLC2000: a new approach to global land cover mapping from Earth observation data. Int. J. Remote Sens. 26, 1959–1977.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.

Blanco, P.D., Colditz, R.R., Saldaña, G.L., Hardtke, L.A., Llamas, R.M., Mari, N.A., Fischer, A., Caride, C., Aceñolaza, P.G., del Valle, H.F., Lillo-Saavedra, M., Coronato, F., Opazo, S.A., Morelli, F., Anaya, J.A., Sione, W.F., Zamboni, P., Arroyo, V.B., 2013. A land cover map of Latin America and the Caribbean in the framework of the SERENA project. Remote Sens. Environ. 132, 13–31.

Bontemps, S., Defourny, P., Bogaert, E., Arino, O., Kalogirou, V., Perez, J., 2011. Globcover 2009 Products Description and Validation Reports. Last accessed February 20, 2017. http://due.esrin.esa.int/files/GLOBCOVER2009_Validation_Report_2.2.pdf.

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. Geocarto Int. 26, 341–358.

Boschetti, L., Roy, D.P., Justice, C.O., Humber, M.L., 2015. MODIS–Landsat fusion for large area 30 m burned area mapping. Remote Sens. Environ. 161, 27–42.

Boschetti, L., Stehman, S.V., Roy, D.P., 2016. A stratified random sampling design in space and time for regional to global scale burned area product validation. Remote Sens. Environ. 186, 465–478.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Campagnolo, M.L., Sun, Q., Liu, Y., Schaaf, C., Wang, Z., Román, M.O., 2016. Estimating the effective spatial resolution of the operational BRDF, albedo, and nadir reflectance products from MODIS and VIIRS. Remote Sens. Environ. 175, 52–64.

Chang, J., Hansen, M.C., Pittman, K., Carroll, M., DiMiceli, C., 2007. Corn and soybean mapping in the United States using MODIS time-series data sets. Agron. J. 99, 1654–1664.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., 2015. Global land cover mapping at 30 m resolution: a POK-based operational approach. ISPRS J. Photogramm. Remote Sens. 103, 7–27.

Claverie, M., Vermote, E.F., Franch, B., Masek, J.G., 2015. Evaluation of the Landsat-5 TM and Landsat-7 ETM+ surface reflectance products. Remote Sens. Environ. 169, 390–403.

Colditz, R.R., 2015. An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms. Remote Sens. 7, 9655–9681.

Colditz, R.R., Saldaña, G.L., Maeda, P., Espinoza, J.A., Tovar, C.M., Hernández, A.V., Benítez, C.Z., López, I.C., Ressl, R., 2012. Generation and analysis of the 2005 land cover map for Mexico using 250 m MODIS data. Remote Sens. Environ. 123, 541–552.

DeFries, R., Townshend, J., 1994. NDVI-derived land cover classifications at a global scale. Int. J. Remote Sens. 15, 3567–3586.

DeFries, R., Hansen, M., Townshend, J., 1995. Global discrimination of land cover types from metrics derived from AVHRR pathfinder data. Remote Sens. Environ. 54, 209–222.

Dieye, A., Roy, D.P., Hanan, N., Liu, S., Hansen, M., Toure, A., 2011. Sensitivity analysis of the GEMS soil organic carbon model to land cover land use classification uncertainties under different climate scenarios in Senegal. Biogeosci. Discuss. 8, 6589.

Doraiswamy, P.C., Hatfield, J.L., Jackson, T.J., Akhmedov, B., Prueger, J., Stern, A., 2004. Crop condition and yield simulations using Landsat and MODIS. Remote Sens. Environ. 92 (4), 548–559.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens. Environ. 120, 25–36.

Egorov, A., Hansen, M., Roy, D.P., Kommareddy, A., Potapov, P., 2015. Image interpretation-guided supervised classification using nested segmentation. Remote Sens. Environ. 165, 135–147.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201.

Foody, G.M., Mathur, A., 2004. Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. Remote Sens. Environ. 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 Sens. Environ. 103, 179–189.

Friedl, M.A., Muchoney, D., McIver, D., Gao, F., Hodges, J.C.F., Strahler, A.H., 2000. Characterization of North American land cover from NOAA-AVHRR data using the EOS MODIS Land Cover Classification Algorithm. Geophys. Res. Lett. 27 (7), 977–980.

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182.

Gibbes, C., Adhikari, S., Rostant, L., Southworth, J., Qiu, Y., 2010. Application of object based classification and high resolution satellite imagery for savanna ecosystem analysis. Remote Sens. 2, 2748–2772.

Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: a review. ISPRS J. Photogramm. Remote Sens. 116, 55–72.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., Chen, J., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. Int. J. Remote Sens. 34, 2607–2654.

Gray, J., Song, C., 2013. Consistent classification of image time series with automatic adaptive signature generalization. Remote Sens. Environ. 134, 333–341.

Griffiths, P., Hostert, P., Gruebner, O., van der Linden, S., 2010. Mapping megacity growth with multi-sensor data. Remote Sens. Environ. 114, 426–439.

Hansen, M.C., Egorov, A., Roy, D.P., Potapov, P., Ju, J., Turubanova, S., Kommareddy, I., Loveland, T.R., 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 Sens. Lett. 2, 279–288.

Hansen, M., Egorov, A., Potapov, P., Stehman, S., Tyukavina, A., Turubanova, S., Roy, D.P., Goetz, S., Loveland, T., Ju, J., Kommareddy, A., Kovalskyy, V., Forsyth, C., Bentsg, T., 2014. Monitoring conterminous United States (CONUS) land cover change with web-enabled Landsat data (WELD). Remote Sens. Environ. 140, 466–484.

Henderson, R., 1976. Signature extension using the MASC algorithm. IEEE Trans. Geosci. Electron. 14, 34–37.

Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Campbell, L.B., 2016. Mass data processing of time series Landsat imagery: pixels to data products for forest monitoring. Int. J. Digital Earth 9, 1035–1054.

Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2017. Disturbance-informed annual land cover classification with Landsat time series: a 29-year national data cube. Remote Sens. Environ. (in review).

Herold, M., Gardner, M.E., Roberts, D.A., 2003. Spectral resolution requirements for mapping urban areas. IEEE Trans. Geosci. Remote Sens. 41, 1907–1919.

Herold, M., Mayaux, P., Woodcock, C., Baccini, A., Schmullius, C., 2008. Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. Remote Sens. Environ. 112, 2538–2556.

Homer, C., Huang, C., Yang, L., Wylie, B., Coan, M., 2004. Development of a 2001 national land-cover database for the United States. Photogramm. Eng. Remote Sens. 70, 829–840.

Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States—representing a decade of land cover change information. Photogramm. Eng. Remote Sens. 81, 345–354.

Huang, C., Song, K., Kim, S., Townshend, J.R., Davis, P., Masek, J.G., Goward, S.N., 2008. Use of a dark object concept and support vector machines to automate forest cover change analysis. Remote Sens. Environ. 112, 970–985.

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., Koetz, B., 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. Remote Sens. 7, 12356–12379.

Irish, R.R., Barker, J.L., Goward, S.N., Arvidson, T., 2006. Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm. Photogramm. Eng. Remote Sens. 72, 1179–1188.

Jia, K., Liang, S., Wei, X., Zhang, L., Yao, Y., Gao, S., 2014. Automatic land-cover update approach integrating iterative training sample selection and a Markov Random Field model. Remote Sens. Lett. 5, 148–156.

Johnson, D.M., 2013. A 2010 map estimate of annually tilled cropland within the conterminous United States. Agric. Syst. 114, 95–105.

Johnson, D.M., Mueller, R., 2010. The 2009 cropland data layer. Photogramm. Eng. Remote Sens. 76, 1201–1205.

Ju, J., Roy, D.P., 2008. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. Remote Sens. Environ. 112, 1196–1211.

Ju, J., Roy, D.P., Vermote, E., Masek, J., Kovalskyy, V., 2012. Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods. Remote Sens. Environ. 122, 175–184.

Justice, C., Townshend, J., Vermote, E., Masuoka, E., Wolfe, R., Saleous, N., Roy, D.P., Morisette, J., 2002. An overview of MODIS Land data processing and product status. Remote Sens. Environ. 83, 3–15.

Kennedy, R.E., Yang, Z., Braaten, J., Copass, C., Antonova, N., Jordan, C., Nelson, P., 2015. Attribution of disturbance change agent from Landsat time-series in support of habitat monitoring in the Puget Sound region, USA. Remote Sens. Environ. 166, 271–285.

Klotz, M., Kemper, T., Geiß, C., Esch, T., Taubenböck, H., 2016. How good is the map? A multi-scale cross-comparison framework for global settlement layers: evidence from Central Europe. Remote Sens. Environ. 178, 191–212.

Knorn, J., Rabe, A., Radeloff, V.C., Kuemmerle, T., Kozak, J., Hostert, P., 2009. Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. Remote Sens. Environ. 113, 957–964.

Kovalskyy, V., Roy, D.P., 2013. The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30 m Landsat data product generation. Remote Sens. Environ. 130, 280–293.
Lawrence, R.L., Moran, C.J., 2015. The AmericaView classification methods accuracy comparison project: a rigorous approach for model selection. Remote Sens. Environ. 170, 115–120.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.
Lindquist, E., Hansen, M., Roy, D.P., Justice, C.O., 2008. The suitability of decadal image data sets for mapping tropical forest cover change in the Democratic Republic of Congo: implications for the mid-decadal global land survey. Int. J. Remote Sens. 29, 7269–7275.
Loveland, T., Merchant, J., Ohlen, D., Brown, J., 1991. Development of a land-cover characteristics database for the conterminous U.S. Photogramm. Eng. Remote. Sens. 57, 1453–1463.
Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L.W.M.J., Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. Int. J. Remote Sens. 21 (6–7), 1303–1330.
Lu, D., Tian, H., Zhou, G., Ge, H., 2008. Regional mapping of human settlements in southeastern China with multisensor remotely sensed data. Remote Sens. Environ. 112, 3668–3679.
Markham, B.L., Storey, J.C., Williams, D.L., Irons, J.R., 2004. Landsat sensor performance: history and current status. IEEE Trans. Geosci. Remote Sens. 42, 2691–2694.
Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Feng, G., Kutler, J., Teng-Kui, L., 2006. A Landsat surface reflectance dataset for North America, 1990–2000. IEEE Geosci. Remote Sens. Lett. 3, 68–72.
McIver, D.K., Friedl, M.A., 2001. Estimating pixel-scale land cover classification confidence using nonparametric machine learning methods. IEEE Trans. Geosci. Remote Sens. 39, 1959–1968.
Millard, K., Richardson, M., 2015. On the importance of training data sample selection in random forest image classification: a case study in peatland ecosystem mapping. Remote Sens. 7, 8489–8515.
Minter, T.C., 1978. Methods of extending crop signatures from one area to another. Proceedings, the LACIE Symposium, a Technical Description of the Large Area Crop Inventory Experiment (LACIE), October 23–26, 1978, Houston, TX.
Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57.
Ozesmi, S.L., Bauer, M.E., 2002. Satellite remote sensing of wetlands. Wetl. Ecol. Manag. 10 (5), 381–402.
Pataki, D.E., Alig, R., Fung, A., Golubiewski, N., Kennedy, C., McPherson, E., Nowak, D., Pouyat, R., Romero Lankao, P., 2006. Urban ecosystems and the North American carbon cycle. Glob. Chang. Biol. 12, 2092–2102.
Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. Remote Sens. Environ. 187, 156–168.
Plourde, J.D., Pijanowski, B.C., Pekin, B.K., 2013. Evidence for increased monoculture cropping in the Central United States. Agric. Ecosyst. Environ. 165, 50–59.
Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C., Defourny, P., 2014. Automated training sample extraction for global land cover mapping. Remote Sens. 6, 3965–3987.
Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., Roberts, D., 2008. Mapping landcover modifications over large areas: a comparison of machine learning algorithms. Remote Sens. Environ. 112, 2272–2283.
Roy, D.P., Kumar, S., 2017. Multi-year MODIS active fire type classification over the Brazilian tropical moist forest biome. Int. J. Digital Earth 10 (1), 54–84.
Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M., Loveland, T.R., Vermote, E., Zhang, C., 2010a. Web-enabled Landsat data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. Remote Sens. Environ. 114, 35–49.
Roy, D.P., Ju, J., Mbow, C., Frost, P., Loveland, T., 2010b. Accessing free Landsat data via the Internet: Africa's challenge. Remote Sens. Lett. 1, 111–117.
Roy, D.P., Wulder, M., Loveland, T., Woodcock, C., Allen, R., Anderson, M., Helder, D., Irons, J., Johnson, D., Kennedy, R., Scambos, T., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymbumer, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014a. Landsat-8: science and product vision for terrestrial global change research. Remote Sens. Environ. 145, 154–172.
Roy, D.P., Qin, Y., Kovalskyy, V., Vermote, E.F., Ju, J., Egorov, A., Hansen, M.C., Kommareddy, I., Yan, L., 2014b. Conterminous United States demonstration and characterization of MODIS-based Landsat ETM+ atmospheric correction. Remote Sens. Environ. 140, 433–449.
Roy, D.P., Zhang, H.K., Ju, J., Gomez-Dans, J., Lewis, P., Schaaf, C., Sun, Q., Li, J., Huang, H., Kovalskyy, V., 2016a. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. Remote Sens. Environ. 176, 255–271.
Roy, D.P., Li, J., Zhang, H.K., Yan, L., 2016b. Best practices for the reprojection and resampling of Sentinel-2 Multi Spectral Instrument Level 1C data. Remote Sens. Lett. 7, 1023–1032.
Schmidt, M., Pringle, M., Devadas, R., Denham, R., Tindall, D., 2016. A framework for large-area mapping of past and present cropping activity using seasonal Landsat images and time series metrics. Remote Sens. 8, 312.
Schneider, A., Friedl, M.A., Potere, D., 2010. Mapping global urban areas using MODIS 500-m data: new methods and datasets based on 'urban ecoregions'. Remote Sens. Environ. 114, 1733–1746.
Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C., Townshend, J.R., 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. Int. J. Digital Earth 6, 427–448.
Sheng, Y., Song, C., Wang, J., Lyons, E.A., Knox, B.R., Cox, J.S., Gao, F., 2016. Representative lake water extent mapping at continental scales using multi-temporal Landsat-8 imagery. Remote Sens. Environ. 185, 129–141.
Sleeter, B.M., Sohl, T.L., Loveland, T.R., Auch, R.F., Acevedo, W., Drummond, M.A., Sayler, K.L., Stehman, S.V., 2013. Land-cover change in the conterminous United States from 1973 to 2000. Glob. Environ. Chang. 23, 733–748.
Small, C., 2005. A global analysis of urban reflectance. Int. J. Remote Sens. 26, 661–681.
Stehman, S.V., 2001. Statistical rigor and practical utility in thematic map accuracy assessment. Photogramm. Eng. Remote Sens. 67, 727–734.
Stehman, S.V., Wickham, J.D., Wade, T.G., Smith, J.H., 2008. Designing a multi-objective, multi-support accuracy assessment of the 2001 National Land Cover Data (NLCD 2001) of the conterminous United States. Photogramm. Eng. Remote Sens. 74, 1561–1571.
Story, M., Congalton, R.G., 1986. Accuracy assessment—a user's perspective. Photogramm. Eng. Remote. Sens. 52, 397–399.
Townshend, J.R., Masek, J.G., Huang, C., Vermote, E.F., Gao, F., Channan, S., Sexton, J.O., Feng, M., Narasimhan, R., Kim, D., Song, K., Song, D., Song, X.-P., Noojipady, P., Tan, B., Hansen, M.C., Li, M., Wolfe, R.E., 2012. Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges. Int. J. Digital Earth 5, 373–397.
Turner, B.L., Lambin, E.F., Reenberg, A., 2007. The emergence of land change science for global environmental change and sustainability. Proc. Natl. Acad. Sci. 104, 20666–20671.
Weiss, G.M., Provost, F., 2003. Learning when training data are costly: the effect of class distribution on tree induction. J. Artif. Intell. Res. 19, 315–354.
Wessels, K.J., van den Bergh, F., Roy, D.P., Salmon, B.P., Steenkamp, K.C., MacAlister, B., Swanepoel, D., Jewitt, D., 2016. Rapid land cover map updates using change detection and robust random forest classifiers. Remote Sens. 8, 888.
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 Sens. Environ. 130, 294–304.
Wickham, J., Stehman, S.V., Gass, L., Dewitz, J.A., Sorenson, D.G., Granneman, B.J., Poss, R.V., Baer, L.A., 2017. Thematic accuracy assessment of the 2011 National Land Cover Database (NLCD). Remote Sens. Environ. 191, 328–341.
Wolfe, R.E., Roy, D.P., Vermote, E., 1998. MODIS land data storage, gridding, and compositing methodology: level 2 grid. IEEE Trans. Geosci. Remote Sens. 36, 1324–1338.
Wolfe, R.E., Nishihama, M., Fleig, A.J., Kuyper, J.A., Roy, D.P., Storey, J.C., Patt, F.S., 2002. Achieving sub-pixel geolocation accuracy in support of MODIS land science. Remote Sens. Environ. 83, 31–49.
Woodcock, C.E., Macomber, S.A., Pax-Lenney, M., Cohen, W.B., 2001. Monitoring large areas for forest change using Landsat: generalization across space, time and Landsat sensors. Remote Sens. Environ. 78, 194–203.
Wulder, M.A., White, J.C., Loveland, T.R., Woodcock, C.E., Belward, A.S., Cohen, W.B., Fosnight, E.A., Shaw, J., Masek, J.G., Roy, D.P., 2016. The global Landsat archive: status, consolidation, and direction. Remote Sens. Environ. 185, 271–283.
Xian, G., Homer, C., Fry, J., 2009. Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods. Remote Sens. Environ. 113, 1133–1147.
Yan, L., Roy, D.P., 2015. Improved time series land cover classification by missing-observation-adaptive nonlinear dimensionality reduction. Remote Sens. Environ. 158, 478–491.
Yan, L., Roy, D.P., 2016. Conterminous United States crop field size quantification from multi-temporal Landsat data. Remote Sens. Environ. 172, 67–86.
Yan, L., Roy, D.P., Zhang, H.K., Li, J., Huang, H., 2016. An automated approach for sub-pixel registration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. Remote Sens. 8, 520.
Yang, L., Huang, C., Homer, C.G., Wylie, B.K., Coan, M.J., 2003. An approach for mapping large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial resolution imagery. Can. J. Remote. Sens. 29, 230–240.
Yu, L., Wang, J., Gong, P., 2013. Improving 30 m global land-cover map FROM-GLC with time series MODIS and auxiliary data sets: a segmentation-based approach. Int. J. Remote Sens. 34, 5851–5867.
Zhang, X., Friedl, M.A., Schaaf, C.B., 2006. Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): evaluation of global patterns and comparison with in situ measurements. J. Geophys. Res. Biogeosci. 111, G04017.
Zhang, H.K., Roy, D.P., Kovalskyy, V., 2016. Optimal solar geometry definition for global long-term Landsat time-series bidirectional reflectance normalization. IEEE Trans. Geosci. Remote Sens. 54, 1410–1418.
Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171.