Remote Sensing of Environment 265 (2021) 112646

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Validation of the U.S. Geological Survey's Land Change Monitoring, Assessment and Projection (LCMAP) Collection 1.0 annual land cover products 1985–2017

Stephen V. Stehman [a,*], Bruce W. Pengra [b], Josephine A. Horton [c], Danika F. Wellington [b]

[a] College of Environmental Science and Forestry, State University of New York, Syracuse, NY 13210, USA
[b] KBR, contractor to the U.S. Geological Survey, Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD 57198, USA
[c] Innovate! Inc., contractor to the U.S. Geological Survey EROS Center, Sioux Falls, SD 57198, USA

ARTICLE INFO

Edited by: Marie Weiss

Keywords:
Accuracy assessment
Remote sensing
Environmental monitoring
Landsat
Time series

ABSTRACT

The U.S. Geological Survey Land Change Monitoring, Assessment and Projection (USGS LCMAP) has released a suite of annual land cover and land cover change products for the conterminous United States (CONUS). The accuracy of these products was assessed using an independently collected land cover reference sample dataset produced by analysts interpreting Landsat data, high-resolution aerial photographs, and other ancillary data. The reference sample of nearly 25,000 pixels and the accompanying 33-year time series of annual land cover reference labels allowed for a comprehensive assessment of accuracy of the LCMAP land cover and land cover change products. Overall accuracy (± standard error) for the per-pixel assessment across all years for the eight land cover classes was 82.5% (±0.2%). Overall accuracy was consistent year-to-year within a range of 1.5% but varied regionally with lower accuracy in the eastern United States. User's accuracy (UA) and producer's accuracy (PA) for CONUS ranged from the higher accuracies of Water (UA = 96%, PA = 93%) and Tree Cover (UA = 90%, PA = 83%) to the lower accuracies of Wetland (UA = 69%, PA = 74%) and Barren (UA = 43%, PA = 57%). For a binary change/no change classification, UA of change was 13% (±0.5%) and PA was 16% (±0.6%) for CONUS when agreement was defined as a match by the exact year of change. UA and PA improved to 28% and 34% when agreement was defined as the change being detected by the map and reference data within a ± 2-year window. Change accuracy was higher in the eastern United States compared to the western US. UA was 49% (±0.3) and PA was 54% (±0.3) for the footprint of change (defined as the area experiencing at least one land cover change from 1985 to 2017). For class-specific loss and gain when agreement was defined as an exact year match, UA and PA were generally below 30%, with Tree Cover loss being the most accurately mapped change (UA = 25%, PA = 31%). These accuracy results provide users with information to assess the suitability of LCMAP data and information to guide future research for improving LCMAP products, particularly focusing on the challenges of accurately mapping annual land cover change.

1. Introduction

Land cover and land cover change, driven by increasing human population, climate change, and resource use decisions and policies, have profound effects on ecosystems at local to global scales and could potentially affect the capacity of those ecosystems to support irreplaceable ecosystem services such as food production, water resources, and climate stability (Foley et al., 2005; Pengra, 2012). Mapping and monitoring land cover and land cover change therefore supplies critical information for a range of societal and scientific purposes (Bontemps et al., 2011; Mueller and Seffrin, 2006). Examples of global change products that have focused on monitoring single classes include tree cover (Hansen et al., 2013), water (Pekel et al., 2016), and bare ground (Ying et al., 2017). At a national scale, the MapBiomas project (Souza et al., 2020) of Brazil monitored multiple land cover classes annually from 1985 to 2017. The MapBiomas project is the most similar national scale effort to the land cover monitoring objectives of our study. Hu and Hu (2019) produced annual land cover maps using Landsat imagery for a 5.5 million km^2 region in Central Asia for the time period 2001–2017. In the United States, the National Land Cover Dataset (Homer et al., 2015)

* Corresponding author.
E-mail address: svstehma@syr.edu (S.V. Stehman).

https://doi.org/10.1016/j.rse.2021.112646
Received 11 June 2021; Received in revised form 31 July 2021; Accepted 10 August 2021
Available online 19 August 2021

has mapped change of eight land cover classes at 5-year intervals from 2001 to 2016. Calderón-Loor et al. (2021) describe a similar national monitoring effort in Australia, mapping change from 1985 to 2015, also in 5-year intervals. The CORINE project (Büttner, 2014; Büttner et al., 2017) has produced pan-European maps of land cover and land cover change with 6-year intervals from 2000 to 2018.

The U.S. Geological Survey (USGS) Land Change Monitoring, Assessment and Projection (LCMAP) initiative has released a suite of annual land cover and land change products for the conterminous United States (CONUS) for the time period of 1985–2017 (Brown et al., 2020). The annual temporal resolution and year-to-year classification stability provide an advance in mapping land cover change that allows meaningful comparison of year-to-year change. Annual land cover maps were produced using all clear Landsat observations in the USGS Landsat Collection 1 Analysis Ready Data (ARD) archive (Dwyer et al., 2018), with training data labels for the land cover classification obtained from the National Land Cover Database (NLCD). The LCMAP land cover legend (Appendix A), generally based on the Anderson Level I classification scheme (Anderson et al., 1976), includes the thematic classes Developed, Cropland, Grass/Shrub, Tree Cover, Wetland, Water, Ice/Snow, and Barren. Complete descriptions of the land cover (Primary Land Cover-LCPRI) and land cover change (Annual Land Cover Change-LCACHG) products are provided in the LCMAP product guide (U.S. Geological Survey, 2020a).

An adaptation of the Continuous Change Detection and Classification (CCDC) time series algorithm (Zhu and Woodcock, 2014) was used to create the LCMAP products. The algorithm used the history of reflectance values at each pixel to build models describing temporal patterns in the sequence of observations and predicted future reflectance values for each pixel. A sustained departure from that spectral/temporal pattern which exceeded calculated thresholds was used to identify land surface change. This change in pattern, and assumed associated change in land surface, ended iterative recalculation of that model and began calculation of a new model that updated over successive observations. The coefficients that mathematically described the patterns of reflectance values used for each model provided a spectral/temporal signature for the land cover present during that stable model (Brown et al., 2020; Zhu and Woodcock, 2014). The coefficient values of these spectral/temporal signatures, along with other ancillary layers, were used as inputs to a gradient boosting decision-tree classification algorithm, XGBoost (Chen and Guestrin, 2016). This classification was used to produce the annual land cover maps. For additional details on the algorithm and classification process, see the LCMAP Algorithm Description Document (U.S. Geological Survey, 2020b) and Brown et al. (2020).

We conducted an accuracy assessment of the LCMAP Collection 1.0 land cover and land cover change products with the objective of describing data product quality. Our assessment adheres to the convention of the Committee on Earth Observing Satellites' Land Product Validation Subgroup, which defines validation as "the process of assessing, by independent means, the quality of the data products derived from the system outputs" (Justice et al., 2000). In this article, we document the methodology and report the results of the accuracy assessment of the LCMAP Collection 1.0 land cover (LCPRI) and land cover change (LCACHG) maps (U.S. Geological Survey, 2020c). Describing the accuracy of a long time series of annual land cover data requires expanding the set of accuracy metrics beyond those used in single date assessments or assessments of change involving just one or two change periods (Cohen et al., 2010; Pontius Jr. et al., 2017). Therefore, we estimated additional accuracy measures that take advantage of the full time series of annual reference data for each sample pixel. These measures included accuracy of the footprint of change, accuracy of the frequency of change or "incidents" (Pontius Jr. et al., 2017), and accuracy of temporal trends in area of change. In addition, we also assessed the stability of accuracy over time (Bontemps et al., 2011).

2. Methods

The LCMAP validation protocol followed good practice guidelines (Olofsson et al., 2014) for accuracy assessment and area estimation. Specifically, the reference sample data were collected for a probability sample of pixels (Section 2.1). For each year of the time series, reference land cover labels for each pixel were determined by a team of interpreters with all interpreters following a common response design that included protocols for enhancing and evaluating inter-interpreter consistency (Section 2.2). These reference labels provided the basis for the accuracy estimates. The accuracy estimates (Section 2.3) were computed following protocols of consistent estimation required for a statistically rigorous analysis (Stehman, 2001). Standard errors were estimated and reported to quantify the uncertainty of the estimates of user's, producer's, and overall accuracies.

2.1. Sampling design

We selected a simple random sample of 24,971 Landsat resolution (30 m × 30 m) pixels from all pixels in the CONUS ARD grid system. The original sample consisted of 25,000 pixels, but a revision of the LCMAP target area to eliminate six ARD tiles in the Great Lakes and Atlantic Ocean that had no land area resulted in 29 of the original sample pixels being outside the newly defined map area. We chose this sampling design to prioritize four desirable design criteria: (1) the design should be a probability sampling design; (2) the design should be simple to implement and require simple estimators of accuracy and area; (3) the design should be easy to augment to increase the sample size after the data collection of the initial sample has been completed; and (4) standard errors can be estimated without use of variance approximations. Another motivation for implementing simple random sampling was because reference data collection began well before the LCMAP land cover maps were finished. Therefore, the common practice of implementing stratified sampling using the map being assessed to define the strata was not a viable option. A disadvantage of simple random sampling is that very rare change types will have small sample sizes, so user's and producer's accuracies of very rare classes may not be precisely estimated. These issues of sampling design are revisited in the Discussion (Section 4.2).

2.2. Response design

Trained interpreters determined the reference land cover labels of the sample pixels for each year of the time series using Landsat imagery, high-resolution aerial photography, and other ancillary datasets (Pengra et al., 2020a). The reference data collection was a collaboration between USGS LCMAP and the U.S. Forest Service Landscape Change Monitoring System (LCMS, 2021). To meet the data requirements of both projects, interpretation protocols were collaboratively developed in a Joint Response Design (JRD) document (U.S. Geological Survey, 2019) and a unique implementation of the TimeSync reference data collection tool was created (Cohen et al., 2010; Pengra et al., 2020a). TimeSync provides interpreters with an efficient display of all available Landsat data, including anniversary date images for each year. Analysts can view all Landsat images as well as change the TimeSync display image to any one of the Landsat images for that year. TimeSync also provides a graphical time series of Landsat values for all usable Landsat observations (Fig. 1). For each sample pixel, analysts visually interpreted the area delineated by the pixel and assigned land use, land cover, and change process attribute labels for each year of the time series, and these attributes were then crosswalked to the LCMAP land cover classes. The land cover reference label was determined for each of the 24,971 sample pixels for each of the 33 years of the time series. We defined land cover agreement as a match between the reference annual land cover label and the LCMAP annual land cover label. The response design was implemented prior to the map products being available so analysts provided reference

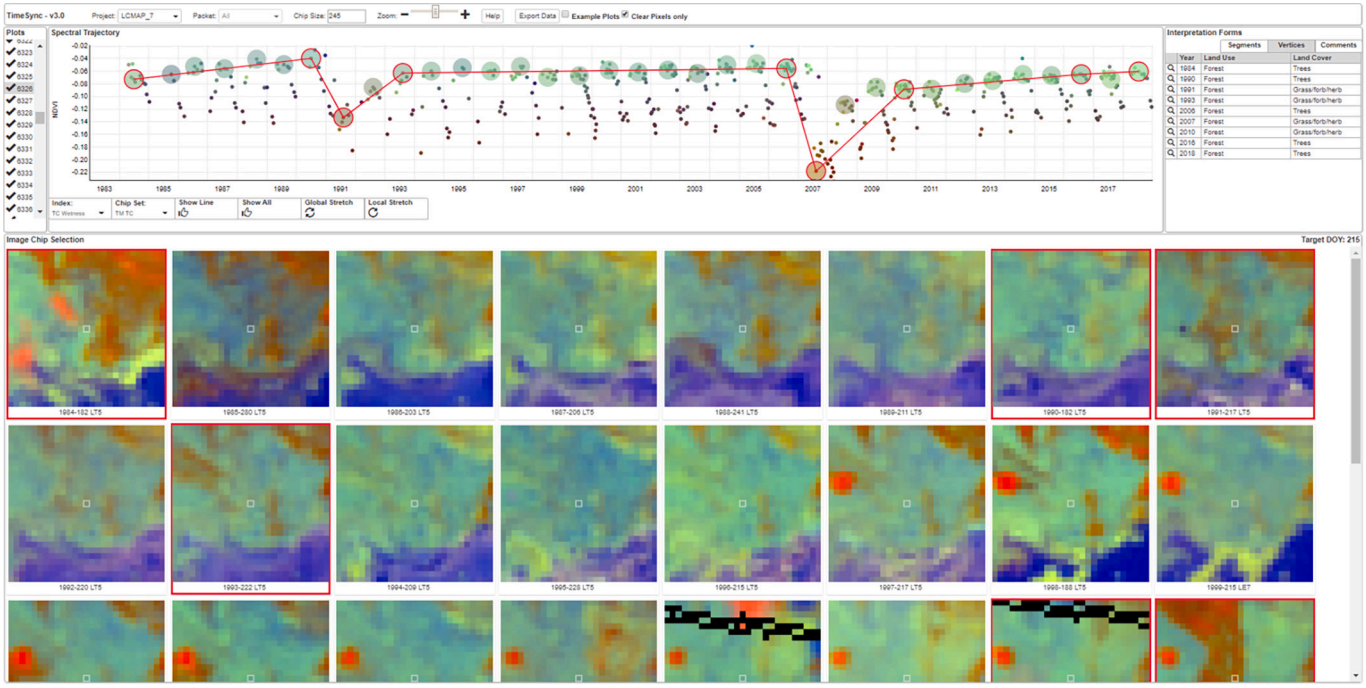

Fig. 1. The TimeSync data collection tool was used to provide efficient display of anniversary date Landsat images (bottom) and graphs of spectral values (6 Landsat bands and 6 derived indices) (top) for all usable observations. TimeSync was also used to record analyst interpretations (top right).

data interpretations with no knowledge of map class labels. Pengra et al. (2020a) document additional details of the response design protocol.

Identifying the occurrence and timing of changes in patterns of Landsat pixel values through time, which correspond to changes from one land cover class to another, is complicated and challenging. The reference data recorded changes in land cover that occurred at any time during a calendar year, and the reference land cover class for that calendar year would be the class interpreted following the change. LCMAP land cover map products were based on a model that was operational on July 1 of each year. This difference between when the map and reference data recorded land cover change could result in both datasets detecting the same event but recording it offset by a year. Consequently, some of the classification errors for map change identified when using the exact year match definition of agreement would be attributable to the offset in how the map and reference data recorded the date of change. To accommodate this offset in reporting year of change and the fact that many land cover changes do not occur abruptly within a single year, we also estimated accuracy with agreement defined as a match within a $\pm$ k year window. For example, for k = 1 the map and reference change would be defined as agreeing if the change was detected by both the map and reference interpretation within a $\pm$ 1 year window.

The spatial unit for the accuracy assessment was a pixel to ensure that the accuracy estimates represented the product delivered to users. In a pixel-based assessment, classification error may be confounded with geolocation error if the reference data and map data are not perfectly aligned. Practitioners implementing spatially explicit analyses using the per-pixel data would experience similar confounding impacts of geolocation error, so the per-pixel accuracy estimates better represent the accuracy that users would encounter in their applications. Blocks and polygons are alternatives to pixels as the assessment units (Stehman and Wickham, 2011; Olofsson et al., 2014). For example, a block assessment unit can be defined by aggregating the land cover data from 3 $\times$ 3 or 5 $\times$ 5 pixel clusters and assigning a new class label based on a majority rule. However, Czaplewski (2003) criticized this approach because accuracy reported for such block units does not address the accuracy of the map delivered to users (i.e., the pixel-based map). Polygon assessment units would be particularly problematic for accuracy assessment of a time series change product simply because polygons could change over time, as for example when a portion of a homogeneous forested polygon undergoes clearing but the remainder of the polygon is still forested. Radoux and Bogaert (2017) pointed out that polygon units are bound to a given partition of the map area and therefore would be difficult to use for change detection accuracy.

2.3. Analysis

The analysis protocol included estimating error matrices and overall, user's, and producer's accuracies. Results are reported for CONUS as a whole and for four large reporting regions (Fig. 2) which aggregate lower-level ecoregions (Omernik and Griffith, 2014). Poststratified estimators (Card, 1982; Stehman, 2013) were used for accuracy estimates for the individual year (annual) land cover maps (e.g., 2010 land cover). Poststratified estimation, a technique for reducing standard errors, was implemented by applying stratified estimator formulas to data from the simple random sample (see Appendix B), where the strata were defined using the map classification. For example, the poststratified estimator of overall accuracy for each year of land cover used the eight land cover classes as the strata.

For accuracy estimates produced from multiple years of data, we treated the sampling design as a one-stage cluster sample in which each pixel was a cluster (the primary sampling unit) and each year of observation for a pixel was a secondary sampling unit. Estimates based on multiple years of data included the error matrix and overall, user's, and producer's accuracies of the annual observations aggregated over all years. Cluster sampling standard error formulas (Stehman, 1997) were used for these accuracy estimates (see Appendix B).

3. Results

We report the accuracy results for the annual land cover maps in Section 3.1 followed by the accuracy of land cover change in Section 3.2. Additional detailed tables of results are available online (Pengra et al., 2020c). Standard errors (SE) of the accuracy estimates are presented in parentheses in the format ($\pm$SE).

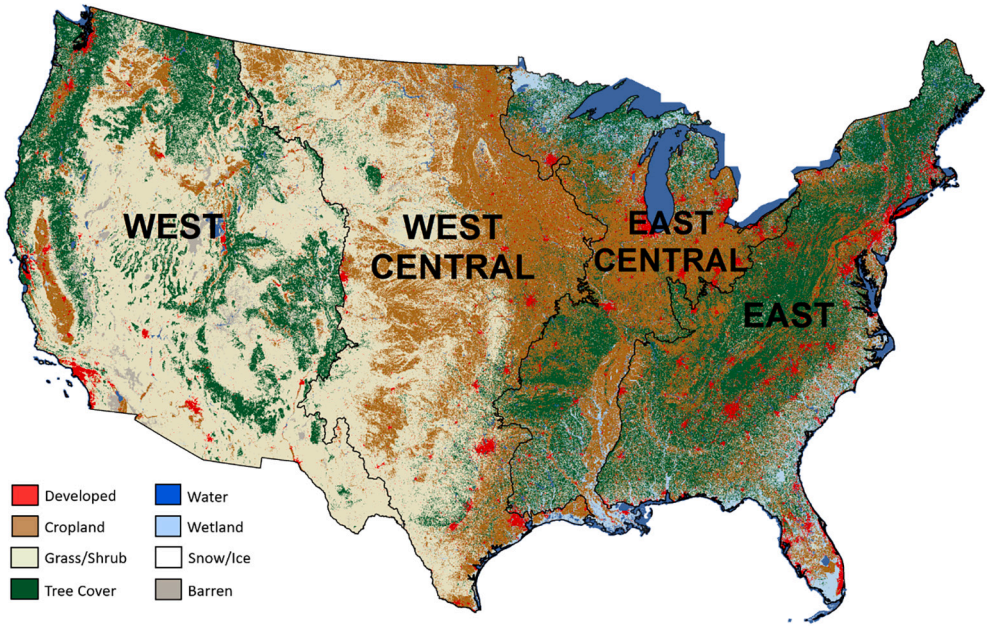

Fig. 2. Spatial boundaries used for regional reporting of LCMAP accuracy overlaid on the LCMAP Collection 1.0 primary land cover (LCPRI) for 2017.

3.1. Land cover accuracy

CONUS overall accuracy was 82.5% (±0.2%) when combining annual land cover accuracy across all years (Table 1). Annual overall accuracy varied within a range of about 1.5% across the time series (1985–2017), peaking at just above 83% (±0.2%) in 1997 (Fig. 3). Annual overall accuracy at the beginning and end of the time series was slightly lower, with the lowest accuracy of 81.5% (±0.2%) occurring in 2017. Overall accuracy was lower in the eastern half of the country compared to the western half (Fig. 3). Across all years, annual overall accuracy fell within a range of 1.4% in the East Central region (least temporally variable) and within a range of 2.2% in the West Central region (most temporally variable). The four regions showed temporal patterns in annual overall accuracy similar to that of CONUS with lower accuracies occurring at the endpoints of the time series (Fig. 3).

CONUS per-class user's (UA) and producer's (PA) accuracies ranged from 43% to 96% (Table 2). Accuracy was greatest for Water with UA of 96% and PA of 93%. Tree Cover and Grass/Shrub were two other accurately mapped classes with UA of 90% and PA of 83% for Tree Cover, and UA of 88% and PA of 80% for Grass/Shrub. Cropland had the largest disparity between UA of 70% and PA of 93% because of substantial commission error to Grass/Shrub (Table 1). Developed had a large disparity between UA of 77% and PA of 63% primarily due to omission errors from Cropland, Grass/Shrub and Tree Cover (Table 1). The largest sources of Developed commission errors were Grass/Shrub and Tree Cover.

Annual UA and PA were generally consistent over time (Figs. 4 and 5), with accuracy of the very rare class Barren the most variable over time. Annual accuracy of Developed showed an upward trend in UA, and, to a lesser extent, a downward trend in PA over time. These trends in accuracy of Developed corresponded with a more rapid increase in Developed area estimated by the reference data relative to the map during the same period. While Developed area increased from 1985 to 2017, the increasing availability of high-resolution data used by the interpreters may have increased the likelihood of resolving features characteristic of Developed land that could not be identified earlier in the time series, thus leading to an increase in the proportion of area of Developed estimated from the sample. These improved ancillary data available to interpreters were not used in the map classification. Consequently, the increasing sensitivity of the reference interpretation to developed features may account for the increasing UA and decreasing PA of the Developed class over time.

UA and PA varied over the four regions (Table 2). Accuracy of Grass/Shrub was much higher in the two western regions, whereas accuracy of Wetland was notably higher in the eastern regions. These regional disparities in accuracy were related to the rarity of the classes in their respective regions. Grass/Shrub was relatively rare in the two eastern regions, comprising about 10% of the area (based on the reference

Table 1
Error matrix for CONUS annual land cover (all years) where cell entries represent percent of CONUS area.

Map	Developed	Cropland	Grass/Shrub	Tree Cover	Water	Wetland	Snow/Ice	Barren	Total	UA (SE)	*n*
Developed	**3.000**	0.139	0.321	0.377	0.024	0.035	0.000	0.001	3.896	77 (1.2)	32,102
Cropland	0.918	**16.527**	5.061	0.799	0.027	0.368	0.000	0.003	23.702	70 (0.6)	195,283
Grass/Shrub	0.368	0.757	**30.649**	2.599	0.045	0.229	0.000	0.332	34.980	88 (0.3)	288,197
Tree Cover	0.340	0.143	1.414	**23.387**	0.049	0.579	0.000	0.006	25.917	90 (0.3)	213,531
Water	0.013	0.008	0.048	0.024	**4.788**	0.067	0.000	0.020	4.968	96 (0.5)	40,932
Wetland	0.062	0.129	0.361	0.944	0.172	**3.688**	0.000	0.001	5.357	69 (1.3)	44,136
Snow/Ice	0.000	0.000	0.004	0.004	0.000	0.004	**0.012**	0.004	0.028	43 (18.7)	231
Barren	0.072	0.005	0.501	0.013	0.056	0.012	0.000	**0.492**	1.151	43 (2.8)	9485
Total	4.772	17.707	38.358	28.149	5.162	4.981	0.012	0.859	100.00		
PA (SE)	63 (1.3)	93 (0.3)	80 (0.4)	83 (0.4)	93 (0.7)	74 (1.2)	100 (0)	57 (3.2)			
n	39,319	145,886	316,027	231,916	42,530	41,042	99	7078			

Overall accuracy was 82.5% (±0.2%). Standard errors for user's accuracy (UA) and producer's accuracy (PA) are shown in parentheses and *n* is the number of sample observations for each row and column. Correct classifications on the diagonal are shown in bold.

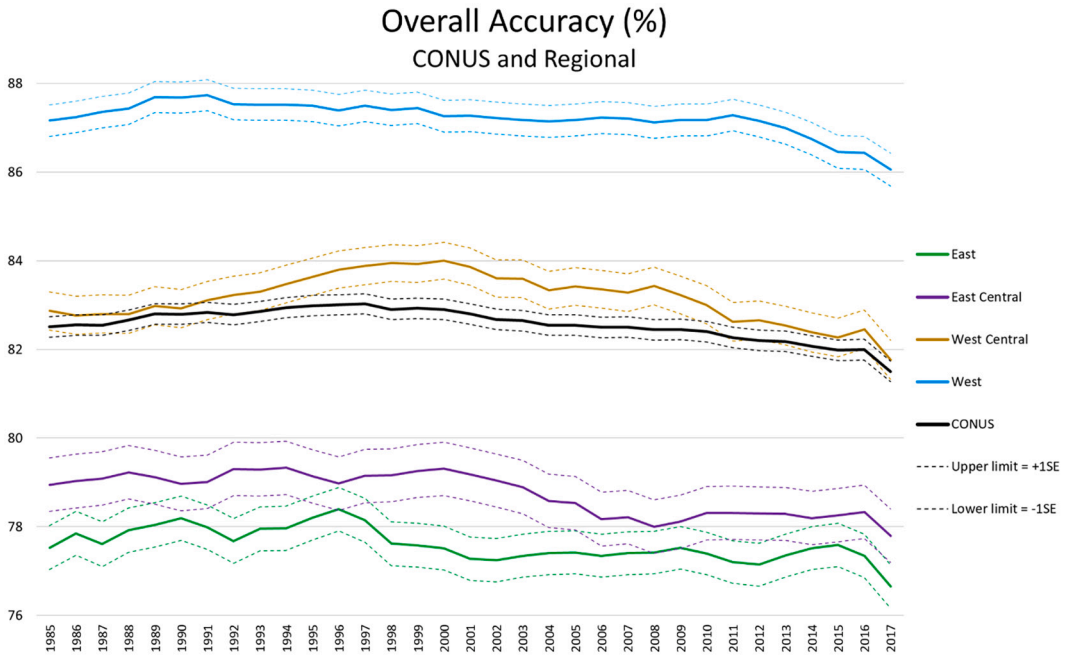

Fig. 3. LCMAP Collection 1.0 overall accuracy of land cover by year (1985–2017).

Table 2
LCMAP user's and producer's accuracies (%) for annual land cover (all years) by CONUS and region (missing values for Snow/Ice are due to having no sample pixels in those regions).

Class	User's accuracy					Producer's accuracy				
	CONUS	East	East Central	West Central	West	CONUS	East	East Central	West Central	West
Developed	77	78	79	73	77	63	64	54	67	69
Cropland	70	47	68	79	72	93	94	96	92	94
Grass/Shrub	88	36	48	89	90	80	15	11	84	91
Tree Cover	90	92	88	88	90	83	86	85	59	84
Water	96	97	98	90	95	93	91	97	84	89
Wetland	69	74	75	39	52	74	83	76	50	44
Snow/Ice	43	–	–	–	43	100	–	–	–	100
Barren	43	0	0	16	51	57	4	0	86	57

classification), but Grass/Shrub was much more common in the western regions, representing 50–60% of the area. Wetlands were more common in the eastern regions, covering about 10% of the area, compared to the western regions where Wetlands covered less than 2% of the area. Classes with low region-specific accuracies included Cropland in the East (UA = 47%), Developed in the East Central (PA = 54%), and Tree Cover in the West Central (PA = 59%). The region-specific error matrices presented in Tables 3–6 provide detailed information regarding omission and commission errors within each of the four regions.

Non-site-specific accuracy is the difference between the map percent area and the reference percent area of each land cover class or land cover change class. This accuracy measure is also called "map bias" because it represents the bias of the area of a class obtained by pixel counting (i.e., summing the area over all pixels that are mapped as the target class) (Olofsson et al., 2014). The largest map bias occurred for Cropland, where the map overestimated the reference area by about 6% for CONUS and by about 10% for the two eastern regions (Table 7). The overestimation of Cropland area resulted in underestimation of area for most other classes. Grass/Shrub area was underestimated by the map, with the magnitude of the underestimation greatest in the two eastern regions. Tree Cover area was underestimated in all four regions, with maximum underestimation of 3.29% occurring in the East region. Developed area was also underestimated by the map in all four regions with the two eastern regions showing a greater degree of underestimation than the two western regions.

3.2. Land cover change accuracy

For binary change/no change and agreement defined as a match by the exact year of change, the CONUS change class had UA of 13% (±0.5%) and PA of 16% (±0.6%) (Table 8). When change agreement was defined as a match between the map and reference data conversion occurring within ±1 year of the nominal year, UA for the change class increased to 24% (±0.7%) and PA increased to 29% (±0.7%) (Table 9). Each year added to the temporal window used to define agreement of the map and reference change increased UA and PA of change by about 4% for CONUS. The two eastern regions had considerably higher UA of change compared to the two western regions, and PA of change was greater in the East and West regions compared to the East Central and West Central regions (Table 9). Map bias of annual change was positive for all regions except the East Central indicating that the map area was greater than the reference area, and the two western regions had the largest overestimates of annual change (Table 9).

For the footprint of change (i.e., pixels for which at least one land cover change occurred during 1985–2017), UA for change was 49% and PA was 54% for CONUS (Table 10). UA for the footprint of change was much greater in the two eastern regions compared to the two western regions, and PA was greater in the East and West regions relative to the

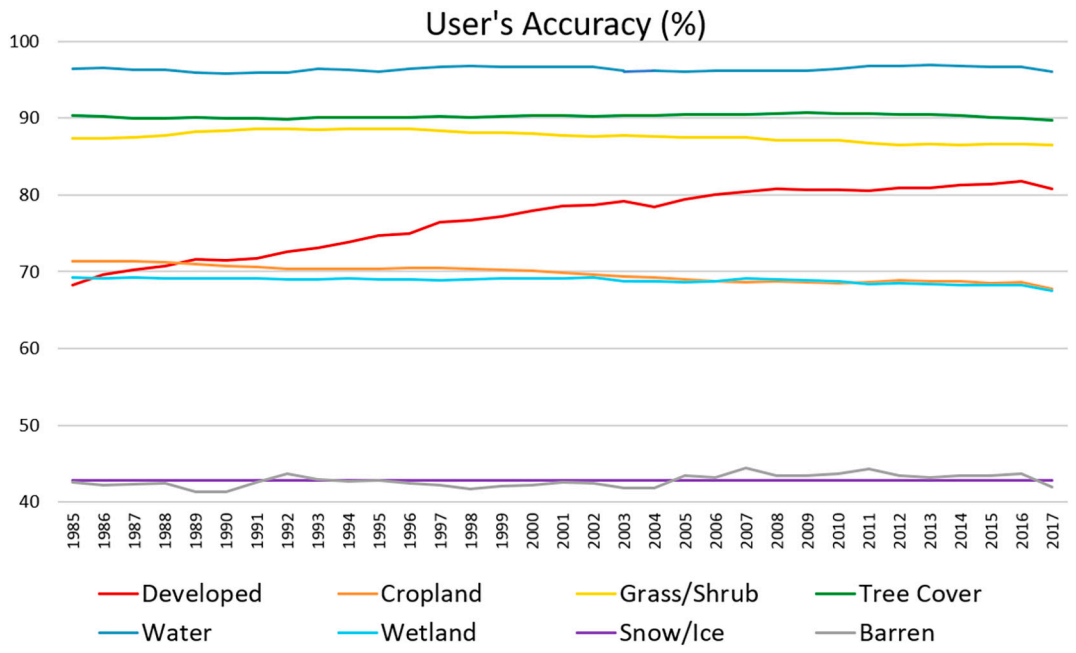

Fig. 4. User's accuracy of land cover by year (1985–2017) for CONUS.

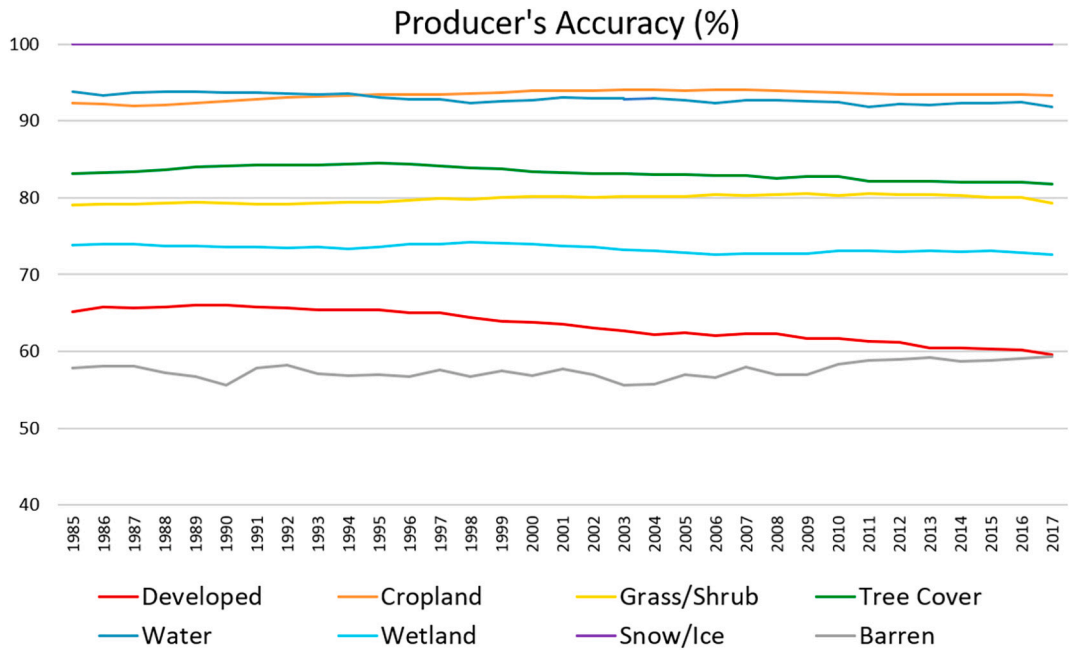

Fig. 5. Producer's accuracy of land cover by year (1985–2017) for CONUS.

East Central and West Central regions. The estimated percent area of the footprint of change based on the reference classification was 11.77% (±0.20) of CONUS, compared to the sample estimated map footprint of change of 12.88% (±0.21) so the map area of the footprint of change exceeded the reference area by an estimated 1.11% at CONUS scale. The map census value for the footprint of change was 13.08%, so was within one standard error of the sample based estimated map area of 12.88%. UA exceeded PA for the footprint of change in the two eastern regions whereas PA exceeded UA in the two western regions. This resulted in the map underestimating the footprint of change in the two eastern regions (due to more omission errors relative to commission errors) and overestimating the footprint of change in the two western regions.

For the frequency of change per pixel (i.e., the number of land cover changes for each pixel), the correlation between the map and reference frequency was 0.45 for CONUS. Regionally, the correlations for frequency of change were greater in the East (0.57) and East Central (0.51) compared to the West (0.39) and West Central (0.27), consistent with the general finding that accuracy of change was greater in the eastern United States (Table 9). In terms of the trend in annual percent area of change, the map time series differed from the reference time series except for the West region (Fig. 6). Although in the West region the temporal trend of the percent area of annual map change mimicked the temporal trend of the reference change area, the map change area was consistently higher than the reference change area. For CONUS and the three regions excluding the West, the map indicated a downward trend in area of change whereas the reference trend was upward in the early

Table 3
Error matrix for East region annual land cover (all years) where cell entries represent percent of area of this region.

Map	Developed	Cropland	Grass/Shrub	Tree Cover	Water	Wetland	Snow/Ice	Barren	Total	UA (SE)	*n*
Developed	**5.788**	0.120	0.255	1.134	0.044	0.107	0.000	0.002	7.449	78 (1.9)	13,394
Cropland	1.626	**9.309**	7.015	1.636	0.036	0.167	0.000	0.001	19.790	47 (1.4)	35,583
Grass/Shrub	0.313	0.135	**1.656**	2.290	0.011	0.162	0.000	0.003	4.571	36 (1.8)	8218
Tree Cover	0.937	0.291	1.458	**45.138**	0.063	1.135	0.000	0.005	49.028	92 (0.5)	88,154
Water	0.021	0.003	0.044	0.081	**7.348**	0.099	0.000	0.002	7.599	97 (0.8)	13,664
Wetland	0.172	0.059	0.205	1.989	0.447	**8.273**	0.000	0.000	11.145	74 (1.7)	20,039
Snow/Ice	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	–	0
Barren	0.143	0.011	0.128	0.047	0.087	0.000	0.000	**0.001**	0.418	0 (0.1)	751
Total	9.002	9.928	10.762	52.316	8.036	9.944	0.000	0.013	100.00		
PA (SE)	64 (2.0)	94 (0.9)	15 (1.0)	86 (0.5)	91 (1.3)	83 (1.6)	–	4 (4.1)			
n	16,185	17,851	19,350	94,065	14,449	17,879	0	24			

Overall accuracy was 77.5% (±0.5%). Standard errors for UA and PA are shown in parentheses and *n* is the number of sample observations for each row and column. Correct classifications on the diagonal are shown in bold.

Table 4
Error matrix for East Central region annual land cover (all years) where cell entries represent percent of area of this region.

Map	Developed	Cropland	Grass/Shrub	Tree Cover	Water	Wetland	Snow/Ice	Barren	Total	UA (SE)	*n*
Developed	**3.775**	0.248	0.229	0.440	0.038	0.031	0.000	0.001	4.761	79 (2.5)	6596
Cropland	2.320	**25.724**	7.397	1.379	0.019	0.841	0.000	0.001	37.680	68 (1.1)	52,200
Grass/Shrub	0.162	0.144	**1.108**	0.840	0.000	0.041	0.000	0.003	2.298	48 (3.5)	3184
Tree Cover	0.546	0.425	1.159	**26.659**	0.122	1.541	0.000	0.001	30.453	88 (0.8)	42,188
Water	0.014	0.031	0.045	0.001	**12.894**	0.183	0.000	0.000	13.167	98 (0.5)	18,241
Wetland	0.069	0.271	0.278	2.030	0.177	**8.604**	0.000	0.000	11.428	75 (2.0)	15,832
Snow/Ice	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	–	0
Barren	0.110	0.001	0.045	0.000	0.043	0.012	0.000	0.000	0.212	0 (0.0)	293
Total	6.995	26.844	10.261	31.350	13.291	11.254	0.000	0.006	100.00		
PA (SE)	54 (2.7)	96 (0.6)	11 (1.1)	85 (0.9)	97 (0.7)	76 (1.9)	–	0 (0.0)			
n	9690	37,188	14,215	43,430	18,413	15,590	0	8			

Overall accuracy was 78.8% (±0.6%). Standard errors for UA and PA are shown in parentheses, and *n* is the number of sample observations for each row and column. Correct classifications on the diagonal are shown in bold.

Table 5
Error matrix for West Central region (all years) where cell entries represent percent of area of this region.

Map	Developed	Cropland	Grass/Shrub	Tree Cover	Water	Wetland	Snow/Ice	Barren	Total	UA (SE)	*n*
Developed	**2.107**	0.199	0.454	0.103	0.012	0.005	0.000	0.000	2.882	73 (2.8)	6856
Cropland	0.391	**30.701**	6.652	0.662	0.028	0.451	0.000	0.000	38.885	79 (0.7)	92,520
Grass/Shrub	0.552	2.209	**43.197**	2.134	0.092	0.348	0.000	0.001	48.534	89 (0.5)	115,478
Tree Cover	0.052	0.026	0.495	**4.880**	0.016	0.083	0.000	0.000	5.552	88 (1.4)	13,209
Water	0.014	0.002	0.076	0.023	**1.379**	0.033	0.000	0.005	1.532	90 (2.3)	3644
Wetland	0.020	0.239	0.654	0.443	0.074	**0.928**	0.000	0.000	2.357	39 (3.6)	5607
Snow/Ice	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	–	0
Barren	0.027	0.000	0.125	0.000	0.044	0.020	0.000	**0.042**	0.259	16 (7.4)	616
Total	3.164	33.376	51.653	8.244	1.645	1.869	0.000	0.049	100.00		
PA (SE)	67 (2.8)	92 (0.5)	84 (0.5)	59 (1.9)	84 (3.0)	50 (4.1)	–	86 (7.7)			
n	7528	79,412	122,898	19,616	3914	4446	0	116			

Overall accuracy was 83.2% (±0.4%). Standard errors for UA and PA are shown in parentheses, and *n* is the number of sample observations for each row and column. Correct classifications on the diagonal are shown in bold.

Table 6
Error matrix for West region annual land cover (all years) where cell entries represent percent of area of this region.

Map	Developed	Cropland	Grass/Shrub	Tree Cover	Water	Wetland	Snow/Ice	Barren	Total	UA (SE)	*n*
Developed	**1.519**	0.041	0.296	0.080	0.013	0.014	0.000	0.000	1.964	77 (2.7)	5256
Cropland	0.183	**4.014**	1.124	0.059	0.025	0.184	0.000	0.007	5.597	72 (1.9)	14,980
Grass/Shrub	0.349	0.200	**54.263**	4.132	0.050	0.265	0.000	1.017	60.276	90 (0.4)	161,317
Tree Cover	0.088	0.001	2.333	**23.534**	0.033	0.147	0.000	0.013	26.148	90 (0.6)	69,980
Water	0.005	0.004	0.025	0.000	**1.904**	0.016	0.000	0.056	2.011	95 (1.3)	5383
Wetland	0.022	0.005	0.247	0.126	0.072	**0.517**	0.000	0.004	0.993	52 (5.4)	2658
Snow/Ice	0.000	0.000	0.012	0.012	0.000	0.012	**0.037**	0.012	0.086	43 (18.7)	231
Barren	0.044	0.007	1.321	0.008	0.053	0.013	0.000	**1.478**	2.924	51 (3.1)	7825
Total	2.211	4.273	59.621	27.951	2.150	1.168	0.037	2.589	100.00		
PA (SE)	69 (3.0)	94 (1.0)	91 (0.4)	84 (0.7)	89 (2.2)	44 (4.9)	100 (0)	57 (3.3)			
n	5916	11,435	159,564	74,805	5754	3127	99	6930			

Overall accuracy was 87.3% (±0.3%). Standard errors for UA and PA are shown in parentheses, and *n* is the number of sample observations for each row and column. Correct classifications on the diagonal are shown in bold.

Table 7
Non-site-specific accuracy (MapBias) quantifying the difference between map and reference percent area by class where map and reference areas are obtained from the row and column totals of Tables 1 and 3–6 (classes listed in order of differences in CONUS percent area).

Class	CONUS			East			East Central			West Central			West		
	Map	Ref	Map Bias	Map	Ref	Map Bias	Map	Ref	Map Bias	Map	Ref	Map Bias	Map	Ref	Map Bias
Grass/Shrub	34.98	38.36	−3.38	4.57	10.76	−6.19	2.30	10.26	−7.96	48.53	51.65	−3.12	60.28	59.62	0.66
Tree Cover	25.92	28.15	−2.23	49.03	52.32	−3.29	30.45	31.35	−0.90	5.55	8.24	−2.69	26.15	27.95	−1.80
Developed	3.90	4.77	−0.87	7.45	9.00	−1.55	4.76	7.00	−2.24	2.88	3.16	−0.28	1.96	2.21	−0.25
Water	4.97	5.16	−0.19	7.60	8.04	−0.44	13.17	13.29	−0.12	1.53	1.65	−0.12	2.01	2.15	−0.14
Snow/Ice	0.03	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.04	0.05
Barren	1.15	0.86	0.29	0.42	0.01	0.41	0.21	0.01	0.20	0.26	0.05	0.21	2.92	2.59	0.33
Wetland	5.36	4.98	0.38	11.15	9.94	1.21	11.43	11.25	0.18	2.36	1.87	0.49	0.99	1.17	−0.18
Cropland	23.70	17.71	5.99	19.79	9.93	9.86	37.68	26.84	10.84	38.89	33.38	5.51	5.60	4.27	1.33

Table 8
Error matrix for annual change/no change (all years) with cell entries representing percent of area of CONUS and agreement defined as a match in the nominal year of the change.

	Reference		Total	UA (SE)	*n*
	No change	Change			
Map					
No change	**98.834**	0.475	99.31	100 (0)	793,412
Change	0.597	**0.093**	0.69	13 (0.5)	5514
Total	99.432	0.568			
PA (SE)	99 (0.0)	16 (0.6)			
n	794,385	4541			

Overall accuracy was 98.9% (±0.02%). Standard errors for UA and PA are shown in parentheses, and *n* is the number of sample observations for each row and column. Correct classifications on the diagonal are shown in bold.

years of the time series, and the reference annual change area tended to be greater than the map annual change area from about 1990–2000. From 2005 onwards for CONUS and the West Central region, the map annual change area trended strongly upwards whereas the reference change area showed a slight downward trend. The East and East Central regions both had relatively stable annual map changes from 2005 to 2017, whereas the reference annual changes showed more dynamic variation in the annual percent area of change during this time period.

Accuracy of annual class-specific land cover conversions with exact year agreement was greater than 25% for UA of Water gain, Water loss, and Tree Cover loss, and for PA of Tree Cover loss (Table 11). When agreement was defined as detecting the change within a ± 2 year window, PA of Tree Cover loss improved to 53% and UA of Water loss and UA of Tree Cover loss both exceeded 40%. In general, and based on the ±2 year agreement window, the most accurate results occurred for Water loss and gain, Tree Cover loss, and Grass/Shrub gain. The annual areas of the individual change types were exceedingly small, with the largest mapped annual changes observed being 0.305% for Grass/Shrub gain and 0.284% for Grass/Shrub loss (Table 11). The full 64 × 64 error matrix showing all 56 possible individual from-to changes and all 8 individual no change classes is reported in Pengra et al. (2020c).

4. Discussion

In the Discussion, we focus primarily on issues of accuracy assessment methodology related to the three component protocols of accuracy assessment, the sampling design, response design, and analysis (Stehman and Czaplewski, 1998). The LCMAP methodology extended the response design and analysis protocols beyond conventional practice (Section 4.1), while paradoxically, the sampling design (simple random) was less complex than the stratified sampling designs conventionally implemented. We discuss some of the unresolved issues regarding how to effectively incorporate stratification to accommodate the many dimensions of accuracy of interest for annual land cover monitoring (Section 4.2). We then continue the discussion of sampling design as it relates to the objectives of map accuracy assessment in the context of design-based versus model-based statistical inference (Section 4.3). In Section 4.4, LCMAP land cover accuracy is compared with other multi-year mapping studies to demonstrate that LCMAP achieved accuracy comparable to these other projects and that commonality exists among

Table 9
User's and producer's accuracies for the change class from the annual binary change/no change classification (all years) with allowances for the change to be detected within ±k years, where k = 1, 2, 3, or 5.

Region	User's accuracy					Producer's accuracy					Area of change (%)		
	Exact	±1	±2	±3	±5	Exact	±1	±2	±3	±5	Map	Ref	MapBias
CONUS	13	24	28	31	35	16	29	34	38	42	0.69	0.57	0.12
East	20	35	42	45	50	20	36	42	46	51	1.03	1.02	0.01
East Central	20	33	39	44	47	15	26	31	34	37	0.49	0.63	−0.14
West Central	6	13	16	18	21	9	19	23	26	30	0.68	0.48	0.20
West	10	17	20	22	25	19	31	36	40	45	0.57	0.31	0.26

The last three columns evaluate the percent area of annual change aggregated over all years.

Table 10
User's and producer's accuracies of change (SE in parentheses) and area comparison for the footprint of change classification by region (the footprint of change is all pixels with at least one land cover change during 1985–2017).

Region	User's accuracy	Producer's accuracy	Map area (%)	Reference area (%)	Map Bias (%)
CONUS	49 (0.9)	54 (0.9)	12.88 (0.21)	11.77 (0.20)	1.11
East	66 (1.5)	59 (1.5)	17.40 (0.51)	19.40 (0.54)	−2.00
East Central	62 (2.5)	43 (2.1)	9.20 (0.45)	13.10 (0.52)	−3.91
West Central	36 (1.5)	44 (1.8)	13.49 (0.40)	10.86 (0.37)	2.63
West	41 (1.6)	68 (2.0)	11.21 (0.35)	6.77 (0.28)	4.44

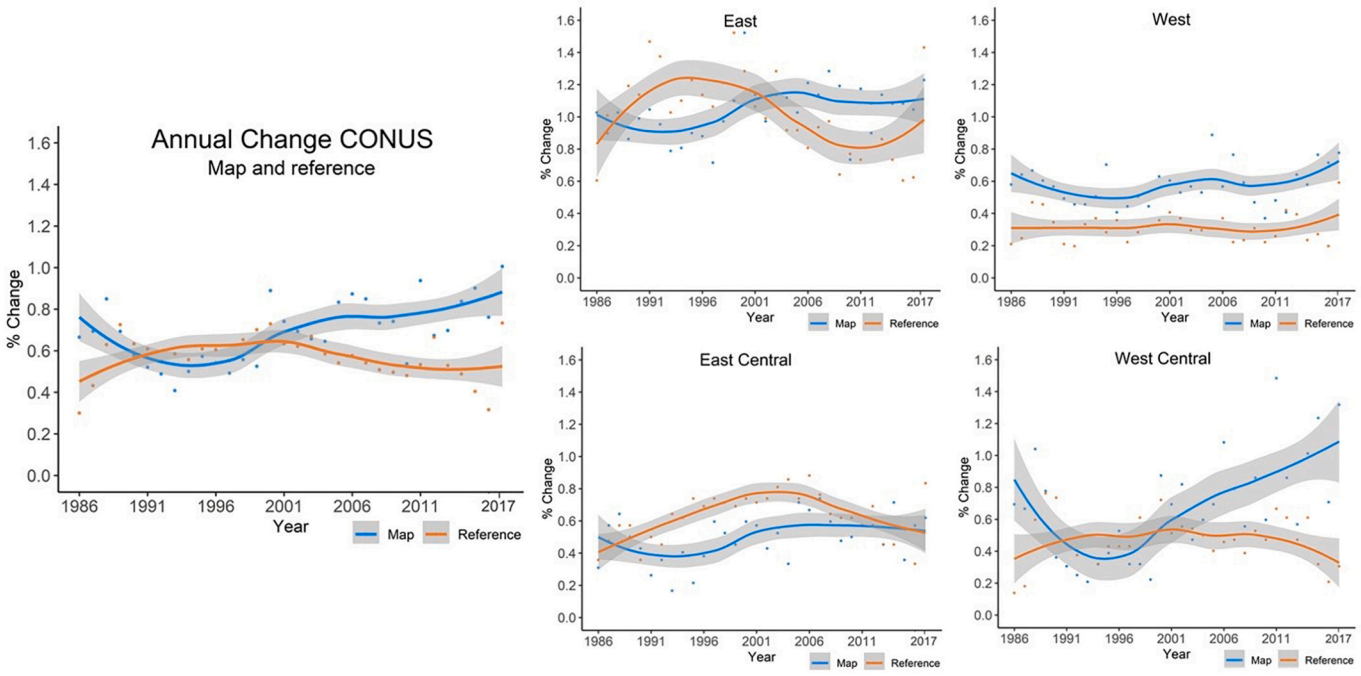

Fig. 6. Comparison of the trend in percent area of annual change for the map and reference classification. The LOESS statistical smoother was applied to determine the trend lines and the gray shading represents a 90% confidence interval). Each dot represents annual percent change area for the region specified.

Table 11
User's (UA), producer's (PA), and non-site-specific (MapBias%) accuracies of annual changes (combined over all years for CONUS).

CONUS class	Exact yr	Exact yr	±2 yrs	±2 yrs	Map area (%)	Reference area (%)	Map Bias (%)
	UA	PA	UA	PA			
Developed gain	12	9	25	19	0.047	0.060	−0.013
Developed loss	0	1	1	6	0.034	0.009	0.025
Cropland gain	3	5	9	15	0.111	0.066	0.045
Cropland loss	5	5	14	15	0.124	0.110	0.014
Grass/Shrub gain	16	20	29	36	0.305	0.243	0.062
Grass/Shrub loss	5	6	16	20	0.284	0.237	0.047
Tree Cover gain	4	5	19	23	0.183	0.151	0.032
Tree Cover loss	25	31	43	53	0.208	0.169	0.039
Water gain	25	16	38	24	0.014	0.022	−0.008
Water loss	27	19	43	30	0.013	0.019	−0.006
Wetland gain	5	2	11	6	0.008	0.015	−0.007
Wetland loss	7	3	12	6	0.007	0.016	−0.009
Barren gain	4	9	8	18	0.022	0.010	0.012
Barren loss	1	3	3	7	0.020	0.009	0.011

projects as to which land cover classes are more challenging to map. Lastly, we briefly discuss potential sources of classification error (Section 4.5).

4.1. LCMAP extensions of response design and analysis protocols

The LCMAP response design included provisions to quantify and enhance agreement among interpreters. Duplicate independent reference class interpretations were obtained for a random subsample of pixels selected from the national sample. These duplicate reference data were used to provide real-time feedback to interpreters about difficult cases, to identify interpreters who may need additional training, and to estimate interpreter agreement (Pengra et al., 2020a). The reference data quality control protocols were carried out during the entire time reference data were collected. To provide transparency of the reference data collection, both the reference class labels as well as the reference sample locations are available (Pengra et al., 2020b) for inspection by those interested in making their own evaluation of the quality of the reference data.

Although the duplicate interpretations furnished information about interpreter consistency, the absence of "gold standard" reference data means that we do not have the information to apply techniques to adjust the accuracy estimates to compensate for reference data error (Foody, 2009, 2010). However, the duplicated interpretations do provide data that would allow incorporating interpreter variance into an estimator of the total variance that included both sampling variance and interpreter variance. McRoberts et al. (2018) provided an estimator of total variance for the estimated proportion of land cover area, although this estimator requires that repeat interpretations are available for the entire sample so it is not applicable for the LCMAP subsample. We are investigating use of a measurement error model approach (Särndal et al., 1992) to estimate total variance for estimated area when duplicate interpretations are available only for a subsample. Estimators of total variance applicable to overall, user's, and producer's accuracies have not been published.

The analysis protocol expanded beyond reporting of conventional accuracy results to provide a richer assessment of accuracy of the LCMAP time series products. These additional analyses included

estimating accuracy of the footprint of change (Table 10), estimating accuracy of trends in mapped percent area of change over time (Fig. 6), estimating accuracy of the number of changes per pixel (i.e., correlation between the map and reference number of changes), and evaluating if annual land cover accuracy varied over time (Figs. 3–5) to assess temporal consistency of accuracy (Bontemps et al., 2011). To summarize the accuracy of the annual land cover maps, we generated error matrices and accuracy estimates by aggregating the annual land cover accuracy data over all years (Tables 1–6). These aggregate estimates required the new development of applying cluster sampling formulas to obtain the standard errors of the aggregate accuracy estimates derived from the annual accuracy data.

4.2. Stratified sampling for accuracy assessment of annual land cover maps

Although stratified sampling is commonly used in accuracy assessment, there are unresolved issues regarding how to effectively implement stratified sampling to accommodate the many estimation objectives that the reference sample is intended to achieve (e.g., annual accuracy of each land cover class, accuracy of different land cover changes, temporal consistency of accuracy, and area of land cover and land cover change). Decisions regarding stratification are dependent on whether the response design specifies obtaining reference class labels for the full time series of annual data (as was the case for LCMAP) or for a single year. Further, decisions regarding strata and sample allocation inherently involve trade-offs related to how well objectives are achieved as different choices will favor precision of some accuracy estimates at the expense of worse precision of other estimates. We address some of these trade-offs in the context of accuracy assessment of annual land cover maps.

An alternative to the LCMAP response design in which a complete time series of annual reference class labels was obtained would be to collect reference class labels for a single year. This single year strategy is equivalent to the "voxel" approach described by Boschetti et al. (2016) for assessing accuracy of burned area in which the sampling unit is viewed as a three-dimensional entity defined by its location in space and the year of observation. In the LCMAP methodology, a voxel sampling unit is a secondary sampling unit because each year of observation of a pixel is considered part of a cluster of years in the full time series of reference class labels observed. These different response design protocols yield alternative options for creating strata to target different accuracy assessment objectives. We examine some of these options using the example of estimating UA of annual Cropland loss.

For the voxel approach, we could construct a stratum consisting of voxels mapped as Cropland loss, and each sample voxel from that stratum would contribute data for estimating UA of Cropland loss. For the full time series response design in which each year represents a secondary sampling unit (SSU) within a pixel (cluster), it is difficult to target the stratification for Cropland loss as combining stratification with cluster sampling is challenging when the stratification is intended to apply to the SSUs (Stehman, 2009, section 6.2). We could assign pixels (clusters) to a stratum that would include all pixels that had mapped Cropland loss for at least one year in the 33-year time series, and then select a sample of pixels from that stratum. This would constitute a stratified, one-stage cluster sample. However, many of the SSUs (years) for a sample pixel from this stratum will not have mapped Cropland loss (and therefore will not contribute to the estimate of UA) because Cropland loss will have occurred once or only a few times in the time series. Alternatively, we could consider two-stage cluster sampling in which we would select a first-stage sample of pixels, assign each SSU (year) from the first-stage pixels to either a Cropland loss stratum or a not Cropland loss stratum, and then select a second-stage sample from the SSUs in the Cropland loss stratum. This two-stage sampling design is more complex to implement and analyze. Further, because ultimately we would collect reference data only for one or a few years from each pixel, the two-stage design is nearly equivalent to a voxel-based approach but with the added complication of having the voxels in clusters.

In this example illustration of the stratification issues, the voxel approach is simpler and allows for a stratification that is more targeted to the single objective of estimating UA of Cropland loss. However, a disadvantage of the voxel approach is that it does not provide the full annual reference data required to evaluate the richer set of change accuracy metrics included in the LCMAP results (see Section 4.1). An additional consideration is whether there may be advantages of better accuracy and efficiency when the reference class interpretations are obtained by analysts having the historical context of the full time series of each pixel versus the isolated year of the time series for the voxel approach. Further exploration of the relative strengths of these approaches to how reference data are obtained, full time series versus single year (voxel), is needed to guide decisions on which approach is better suited for different accuracy assessment objectives and how this decision affects the choice of strata.

In the general perspective of stratification for accuracy assessment, for single date land cover maps it is straightforward to define strata based on the land cover classes and decide stratum-specific sample sizes targeting precision of UA (although this common strategy does not consider precision of producer's accuracy and area estimates). Conversely, for annual land cover maps, the number of potential strata increases dramatically if we want to obtain precise estimates of UA of each land cover class as well as the many land cover change types. For example, Wickham et al. (2017) defined 38 strata in each of two geographic regions in an effort to obtain precise estimates of UA for land cover change mapped at 5-year intervals, 2001–2006–2011. At one end of the spectrum of stratification options, we have the "deep" stratification (numerous strata) such as implemented by Wickham et al. (2017), and at the other end we have the LCMAP simple random sample with no strata. Investigating different options for defining strata within this range of sampling designs is necessary where this evaluation must take into account trade-offs in precision among the many accuracy estimates of interest in an assessment of annual land cover data.

Regardless of the choice of strata and sample size allocated to strata, the stratified estimators of accuracy and area are unbiased. Decisions of which strata and sample allocation to use impact precision of the accuracy estimates. While we would expect to observe different standard errors depending on these decisions, the fact that the estimators are unbiased ensures that the accuracy estimates are still targeting the same parameters and therefore we would not obtain lower or higher accuracies because of the choice of strata and sample allocation. Research is needed to better determine the specific nature of the trade-offs in precision so that strata and sample size allocation can be decided to favor the priority estimates specified for a particular application.

Given the motivation and common use of stratified sampling, it is reasonable to question how well the LCMAP simple random sample performed in terms of precision of accuracy estimates for rare classes. Typically, strata are defined based on the map classes to target the objective of estimating UA and specifying a minimum sample size per stratum (class) to ensure that UA of each class is estimated with a desired level of precision. If the sample size is large, as was the case for LCMAP, adequate precision of accuracy estimates may be still be achieved for rare classes without stratifying the sampling design. For example, if we specify a standard error of 5% as the precision we would like to achieve, then a sample size of 75 for a class that has UA of 70% and a sample size of 51 for a class with UA of 85% would be required to meet that precision goal (see Eq. (5) of Appendix B). For a simple random sample of 25,000 pixels, a class that occupied 0.3% of CONUS area would have an expected sample size of 75 pixels ($0.003 \times 25{,}000$) and a class that occupied 0.2% of CONUS would have an expected sample size of 51. Consequently, the large size of the LCMAP simple random sample would yield satisfactory precision (standard error of 5%) for a class with UA of 70% if that class covered 0.3% of CONUS, and for a class with UA of 85%

if that class covered 0.2% of CONUS. These are just two examples of the many conditions of interest, but these examples illustrate that at CONUS scale, the large sample size will support precise estimates for relatively rare classes.

4.3. Sampling design and statistical inference

For accuracy assessments of long-term, annual land cover data, it remains imperative that the methods adhere to good practice guidance (Olofsson et al., 2014; Stehman and Foody, 2019). Accordingly, we implemented the LCMAP accuracy assessment based on probability sampling and design-based inference so that the reported accuracies satisfied criteria of rigorous inference (i.e., unbiased estimation of accuracy accompanied by standard errors to quantify uncertainty). Accuracy assessments conducted for the objective of describing accuracy have historically implemented probability sampling designs and invoked design-based inference (e.g., Stehman, 1992, 2000). Recent attention to model-based inference in the form of estimating "predictive accuracy" (Ploton et al., 2020; Karasiak et al., 2021) motivates revisiting the distinction between design-based and model-based inference with the sampling design being the focal point of the contrasting views.

In the model-based framework of predictive accuracy, Ploton et al. (2020) and Karasiak et al. (2021) advocate implementing a sampling design in which the validation or test sample pixels must be a sufficient distance from training pixels to be considered independent, where this minimal distance is ascertained based on a measure of spatial autocorrelation. Conversely, design-based inference requires probability sampling in which every element of the region mapped (e.g., a pixel in CONUS) must have a non-zero probability of being included in the sample. The sampling design considerations for predictive accuracy, therefore, are fundamentally different from those of descriptive accuracy. Further, describing accuracy of completed map products such as LCMAP land cover and land cover change is a very different objective from evaluating predictive accuracy of a model or classifier.

In model-based inference, the population is viewed as a process, for example, the classification process generating the mapped land cover. Inference to assess performance of a predictive model such as a land cover classification algorithm is then intended to apply to regions or time periods that are not represented by the training sample data. That is, how well would this classification algorithm perform in another place and at another time? Conversely, the descriptive objective of the LCMAP accuracy assessment characterizes the accuracy of the land cover products for the space (CONUS) and time (1985–2017) actually mapped. That is, the accuracy parameters defined in the design-based framework are quantities representing a real population for which each element of the population is potentially observable (Stehman, 2000). For example, UA of Tree Cover for LCMAP in 1995 is a statement about the LCMAP land cover product for all pixels in CONUS for the year 1995, not a prediction of how accurately the LCMAP procedure would map Tree Cover if applied elsewhere. In design-based inference, all pixels in the mapped region must be eligible to be sampled, including those that were part of, or in close proximity to, the training sample pixels. This is because all pixels are part of the population of interest for the objective of describing the accuracy of map products such as those provided by LCMAP. Consequently, imposing a distance constraint from training sample pixels, as advocated by Ploton et al. (2020) and Karasiak et al. (2021), violates a key tenet of probability sampling required to invoke design-based inference.

Design-based inference represents a statistically valid foundation for rigorous assessment of map accuracy. It is straightforward to demonstrate that estimators of map accuracy are unbiased in this framework (cf. Stehman, 1992), and sampling theory also proves the unbiased property of accuracy estimators in the design-based setting (de Gruijter and ter Braak, 1990; Stehman, 2000). There may be applications in which model-based inference and predictive accuracy are relevant. However, this is not the case for the LCMAP accuracy assessment for which the objective was to describe accuracy of the annual land cover and land cover change products, not to evaluate how well the LCMAP classification methods would perform for different regions outside of CONUS or time periods different from 1985 to 2017. The admonition to implement sampling designs for accuracy assessment that impose spatial distance requirements for the alleged purpose of reducing "the gap between the accuracy statistics given to users and the real quality of the maps produced" (Karasiak et al., 2021) is incorrect because design-based inference invoked for reference data obtained from a probability sampling design is fully justified for the objective of descriptive accuracy that is the intent of projects such as LCMAP.

4.4. Comparison of LCMAP accuracy to other land cover mapping projects

We compared the LCMAP land cover accuracy to accuracy of three other large-scale mapping projects that included a time series of land cover data: (1) the National Land Cover Dataset (NLCD) which mapped eight Anderson Level 1 classes at five-year intervals from 2001 to 2016 (Wickham et al., 2021, 2) the MapBiomas project that generated annual land cover maps for Brazil for the timespan 1985–2017 (Souza et al., 2020); and (3) a mapping study by Hu and Hu (2019) over a region of Central Asia that encompassed Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan, and the Xinjiang region of China, from 2001 to 2017. These accuracy results (Table 12) for the different projects are not strictly comparable in the sense of assessing classification approaches because differences in geography, class definitions, and other factors preclude an unambiguous comparison. Rather, the goal is to illustrate broader tendencies in accuracy among projects with similar mapping objectives.

In general, LCMAP attained land cover accuracy that was broadly consistent with accuracy of these other large-scale studies. All four products had high UA and PA for Forest and Tree Cover classes, as well as for Water, consistent with an expectation that these land cover types are inherently easier to classify by spectral means. The relatively uncommon classes of Wetland and Barren were generally the least accurate within each project. An exception was the high Barren accuracy reported by Hu and Hu (2019), likely aided by the large expanses of desert areas within the arid Central Asia region. Developed and urban classes generally had low PA, which may reflect the challenges of reliably identifying an inherently fragmented land cover class that contains features such as narrow roads that are below the resolution of 30 m Landsat data products. Accuracies among land cover classes that are the approximate equivalent of LCMAP Grass/Shrub were quite variable among projects. Both NLCD and the Central Asia study disaggregated grassland from shrubland, and accuracies of each of these classes may reflect variations in climate, geography, and land use patterns between study regions. Both Wickham et al. (2021) and Souza et al. (2020) noted substantial class confusion among pasture, agriculture, and grassland, whereas NLCD and Hu and Hu (2019) observed misclassification between grassland and shrubland classes. The manner in which spectrally similar and transitional classes are combined in land cover hierarchies can be expected to substantially impact accuracy.

4.5. Potential sources of error

Issues affecting LCMAP land cover accuracy could arise from several sources including violations of the fundamental assumptions underlying the LCMAP CCDC approach, break detection accuracy of the time series segmentation algorithm, errors in the classification training dataset, and Landsat data quantity and quality. The LCMAP modelling approach assumes that the spectral behavior of the land surface through time can be represented with annual harmonic models that are easily separated into discrete periods of time. Therefore, the model is expected to perform better when the short-term spectral variability of the land surface is low, the changes being sought after have a large spectral response, and the

Table 12
Comparison of LCMAP class-specific accuracies of annual land cover with accuracy of other monitoring studies, Central Asia (Hu and Hu, 2019, table 2), MapBiomas (Souza et al., 2020, tables 4 and 5), and NLCD (Wickham et al., 2021).

Hu & Hu, Central Asia	UA	PA	MapBiomas, Brazil	UA	PA	NLCD, CONUS	UA	PA	LCMAP, CONUS	UA	PA
Artificial	63	77	Urban	93	63	Urban	91	65	Developed	77	63
Cultivated	83	85	Agriculture	81	83	Agriculture	89	86	Cropland	70	93
Grassland	92	90	Grassland	50	60	Grassland	65	67	Grass/Shrub	88	80
Shrubland	63	77	Pasture	91	76	Shrubland	67	81			
Forest	95	95	Forest	92	95	Forest	91	80	Tree Cover	90	83
Water	97	96	Water	89	85	Water	92	94	Water	96	93
Wetland	80	80	Wetland	56	53	Wetland	55	90	Wetland	69	74
Snow/Ice	96	95							Snow/Ice	43	100
Barren	97	94				Barren	58	63	Barren	43	57

Land cover class names follow those used for each study and class definitions may not be consistent across studies. Overall accuracy was 90% for Central Asia, 89% for MapBiomas, 79% for NLCD, and 83% for LCMAP.

observational data density is high. All of these factors may be expected to be dependent on the regional geography and vary across different landscapes. For example, the quantity of available cloud- and snow-free Landsat data varies considerably across the CONUS, and is further modified by the presence of swath overlap (or sidelap) zones of Landsat acquisition footprints. How the algorithm responds to these variations in input data contributes to the accuracy of the mapped change. For classification, the selection of training data and accuracy of training data labels may both exert considerable influence on accuracy. Decisions made with respect to the filtering and balance of the pool of training data points may have different effects on accuracy of land cover and accuracy of land cover change. Lastly, aligning the map and reference data to record the year of change in a similar manner would remove a source of error that arose from the asynchronous development of the classification algorithm and response design.

5. Conclusions

Assessing the accuracy of a 33-year annual time series of land cover data creates numerous methodological challenges not encountered for single date and less temporally dense (e.g., 5-year intervals) land cover products. The methods implemented for LCMAP accuracy assessment adhered to good practice protocols (Olofsson et al., 2014) and yielded a diverse array of results documenting different aspects of accuracy of the annual land cover and change products. At CONUS scale, the overall accuracy of annual LCMAP land cover was consistently about 82.5% and UA and PA were 70% or greater for most classes. Although LCMAP annual land cover accuracy remained stable over the long time series, accuracy varied substantially over the four regions (Table 2, Fig. 3). Accuracy also varied among classes, with lower accuracies observed for Developed and Barren, and higher accuracies observed for Tree Cover, Water, and Cropland. In contrast to the consistently high overall accuracy of the annual land cover maps, the CONUS site-specific UA and PA for a binary change/no-change classification were below 20% when the definition of agreement required an exact match of the year of change. UA and PA improved to 28% and 34% when agreement was defined as a match within ±2 years for detecting the change. UA and PA for the footprint of change were 49% and 54% indicating better performance at detecting this feature of change. Accurately mapping annual change over a 30+ year time series is a daunting task, particularly considering that only about 0.6–0.7% of the area of CONUS changed in any given year and that the area of each constituent class-specific change was even smaller.

The LCMAP accuracy assessment identified the spatial, temporal, and thematic structure of classification errors in the LCMAP land cover and land cover change products. These results inform further research and development aimed at improving the accuracy of the map products. The full LCMAP Collection 1.0 product suite contains additional data products that describe the behavior of the spectral harmonic models and provide classification confidence scores that underlie the land cover interpretation. This information will contribute to in-depth investigation of sources of classification error and strategies for improving the change detection algorithm, the training data, and the land-cover classification algorithm to enhance the quality of future LCMAP products.

Credit author statement

BP and JH developed the response design for collecting reference data. SS and DW derived and implemented the statistical analysis. All authors participated in writing and revising the manuscript.

Declaration of Competing Interest

The authors declare no conflict of interest. The reference data collection and map accuracy assessment were carried out independently of the mapping process as the team that conducted the accuracy assessment was minimally involved with creation of the map products.

Acknowledgments

We thank three anonymous reviewers for their very helpful and constructive comments. This work was supported with funding from the U.S. Geological Survey's (USGS) Core Science Systems Mission Area, in part under USGS Contract 140G0121D0001 and by USGS Cooperative Agreement G17AC00237 provided to SUNY ESF. This research was also funded by the U.S. Forest Service (IRDB-LCMSNWFP) Landscape Change Monitoring System, the U.S. Forest Service Information Resources Direction Board and the Northwest Forest Plan. Zhiqiang Yang provided generous assistance to implement TimeSync, Warren Cohen and Sean Healey were instrumental in the interagency reference data collection effort. Alexander Hernandez of Utah State University (USU) provided TimeSync technical assistance and supported the USU student interpreters. We thank the interpreters who collected reference data: Roger Auch, Janis Taylor, Steve Kambly, Daryn Dockter, Alexandra Hernandez, Ellie Leydsman, Mikala Solos, Tommy Thompson, Dalton Newbold, Todd Schroeder and Helen Zhong. Matthew Rigge, Thomas Loveland, and Dingfan Xing provided helpful comments on drafts of the manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. LCMAP land cover class definitions

Developed – Areas of intensive use with much of the land covered with structures (e.g., high density residential, commercial, industrial, or

transportation), or less intensive uses where the land cover matrix includes vegetation, bare ground, and structures (e.g., low density residential, recreational facilities, cemeteries, transportation/utility corridors, etc.), including any land functionally related to the developed or built-up activity.

Cropland – Land in either a vegetated or unvegetated state used in production of food, fiber, and fuels. This includes cultivated and uncultivated croplands, hay lands, orchards, vineyards, and confined livestock operations. Forest plantations are considered as forests or woodlands (Tree Cover class) regardless of the use of the wood products.

Grass/Shrub – Land predominantly covered with shrubs and perennial or annual natural and domesticated grasses (e.g., pasture), forbs, or other forms of herbaceous vegetation. The grass and shrub cover must comprise at least 10% of the area and tree cover is less than 10% of the area.

Tree Cover – Tree-covered land where the tree cover density is greater than 10%. Cleared or harvested trees (i.e., clearcuts) will be mapped according to current cover (e.g., Barren, Grass/Shrub).

Water – Areas covered with water, such as streams, canals, lakes, reservoirs, bays, or oceans.

Wetland – Lands where water saturation is the determining factor in soil characteristics, vegetation types, and animal communities. Wetlands are comprised of mosaics of water, bare soil, and herbaceous or wooded vegetated cover.

Ice/Snow – Land where accumulated snow and ice does not completely melt during the summer period (i.e., perennial ice/snow).

Barren – Land comprised of natural occurrences of soils, sand, or rocks where less than 10% of the area is vegetated.

Appendix B. Accuracy and area estimation for land cover and land cover change

A variety of sample-based estimators were applied depending on the accuracy parameter being estimated. The three general types of estimators can be categorized under the headings simple random sampling, cluster sampling, and poststratified estimation. Simple random sampling formulas were applied to the footprint of change results (Table 10) and the annual user's accuracy estimates displayed in Fig. 4. Cluster sampling formulas were applied to all estimators of accuracy that aggregated annual data (e.g., Tables 1–9 and 11). Poststratified estimation was applied to obtain the CONUS and regional annual overall accuracies shown in Fig. 3 and the annual producer's accuracies shown in Fig. 5.

For simple random sampling, the proportion of area in cell (i, j) of the error matrix (i = row, j = column) is estimated by

$$\widehat{p}_{ij} = \frac{n_{ij}}{m} \tag{1}$$

where n_{ij} = number of sample pixels with map class of row i and reference class of column j, and m = number of sample pixels in the region for which accuracy is being estimated (e.g., m = 24,971 for CONUS). User's accuracy for class i is then estimated by

$$\widehat{U}_i = \frac{n_{ii}}{n_{i\cdot}} \tag{2}$$

where $n_{i\cdot} = \sum_{j=1}^{q} n_{ij}$ is the number of sample pixels mapped as class i, and q = number of classes. Producer's accuracy for class j is estimated by

$$\widehat{P}_j = n_{jj} \Big/ n_{\cdot j} \tag{3}$$

where $n_{\cdot j} = \sum_{i=1}^{q} n_{ij}$ is the number of sample pixels with reference class j. Overall accuracy is estimated by

$$\widehat{O} = \sum_{i=1}^{q} n_{ii} \Big/ m \tag{4}$$

The standard errors for the accuracy estimators for simple random sampling are

$$SE\left(\widehat{U}_i\right) = \sqrt{\frac{\widehat{U}_i\left(1-\widehat{U}_i\right)}{n_{i\cdot}-1}}, \tag{5}$$

$$SE\left(\widehat{P}_j\right) = \sqrt{\frac{\widehat{P}_j\left(1-\widehat{P}_j\right)}{n_{\cdot j}-1}}, \tag{6}$$

and

$$SE\left(\widehat{O}\right) = \sqrt{\frac{\widehat{O}\left(1-\widehat{O}\right)}{m-1}}. \tag{7}$$

For simple random sampling, the estimated proportion of area of class j based on the reference classification (see Table 10) is

$$\widehat{p}_k = n_k \Big/ m \tag{8}$$

where n_k = number of sample pixels of reference class k. The standard error is

$$SE\left(\widehat{p}_k\right) = \sqrt{\frac{\widehat{p}_k\left(1-\widehat{p}_k\right)}{m-1}} \tag{9}$$

For estimates that aggregate data across multiple years (e.g., aggregating annual accuracy over all years), the estimators of accuracy and area are similar to those used for simple random sampling, but the standard errors are estimated using one-stage cluster sampling formulas, with the pixel as the primary sampling unit (m denotes the number of pixels) and the year of observation as the secondary sampling unit within each pixel. The estimator of the proportion of area for a cell of the error matrix is

$$\widehat{p}_{ij} = \frac{n_{ij}}{n} \tag{10}$$

where n_{ij} = number of sample pixels with map class of row i and reference class of column j for all years combined, and n = number of observations over all sample pixels and years (e.g., for CONUS, n is 24,971 pixels times 33 years which is 824,109). User's accuracy for class i is estimated by

$$\widehat{U}_i = \frac{n_{ii}}{n_{i\cdot}} \tag{11}$$

where $n_{i\cdot} = \sum_{j=1}^{q} n_{ij}$ is the number of sample pixels mapped as class i over all years combined. Producer's accuracy for class j is estimated by

$$\widehat{P}_j = n_{jj} \Big/ n_{\cdot j} \tag{12}$$

where $n_{j} = \sum_{i=1}^{q} n_{ij}$ is the number of sample pixels with reference class j over all years combined. Overall accuracy is estimated by

$$\widehat{O} = \sum_{i=1}^{q} n_{ii} \Big/ n \tag{13}$$

The cluster sampling standard errors are (Stehman 1997):

$$SE\left(\widehat{U}_i\right) = \frac{1}{\overline{n}_{i\cdot}} \sqrt{\frac{\sum_{u=1}^{m}\left(n_{ii,u} - \widehat{U}_i n_{i\cdot,u}\right)^2}{m(m-1)}}, \tag{14}$$

$$SE\left(\widehat{P}_j\right) = \frac{1}{\overline{n}_{\cdot j}} \sqrt{\frac{\sum_{u=1}^{m}\left(n_{jj,u} - \widehat{P}_j n_{\cdot j,u}\right)^2}{m(m-1)}}, \tag{15}$$

and

$$SE\left(\widehat{O}\right) = \sqrt{\frac{\sum_{u=1}^{m}(p_u - \overline{p})^2}{m(m-1)}} \tag{16}$$

The cluster sampling standard errors require keeping track of information on a per cluster basis, so the subscript u is used to denote a cluster (pixel). In the standard error formula for user's accuracy,

$$\overline{n}_{i\cdot} = \sum_{u=1}^{m} n_{i\cdot,u} \Big/ m, \tag{17}$$

where $n_{i\cdot,\,u}$ is the total number of all secondary sampling units in map class i for cluster u (i.e., total number of pixels over all years) and $n_{ii,\,u}$ is the total number of secondary units assigned to map class i and reference class i for cluster u (the number of years pixel u had map class i and reference class i). In the standard error formula for producer's accuracy,

$$\overline{n}_{\cdot j} = \sum_{u=1}^{m} n_{\cdot j,u} \Big/ m \tag{18}$$

where $n_{j,\,u}$ is the total number of sample units in reference class j for cluster u. In the standard error formula for overall accuracy,

$$p_u = \sum_{j=1}^{q} \widehat{p}_{jj,u}, \tag{19}$$

where $\widehat{p}_{jj,u}$ is the proportion of pixels with map class j and reference class j for cluster u, and $\overline{p} = \sum_{u=1}^{m} p_u/m$.

Poststratified estimators are used to reduce the standard errors of overall and producer's accuracies. For the poststratified estimator,

$$\widehat{p}_{ij} = W_i \frac{n_{ij}}{n_{i\cdot}}, \tag{20}$$

where W_i is the proportion of area mapped as class i, n_{ij} is the number of sample units that have map class i and reference class j, and $n_{i\cdot}$ is the total number of sample units in map class i. The estimators of producer's accuracy (P_j) of reference class j and overall accuracy (O) for q classes are therefore:

$$\widehat{P}_j = \frac{\widehat{p}_{jj}}{\widehat{p}_{\cdot j}}, \tag{21}$$

and

$$\widehat{O} = \sum_{i=1}^{q} \widehat{p}_{ii}, \tag{22}$$

where $\widehat{p}_{i\cdot}$ is the sum of the estimated area proportions in map class i, and $\widehat{p}_{\cdot j}$ is the sum of the estimated proportions in reference class j. The standard errors of the poststratified estimators are similar to those for a stratified sampling design based on the conditional analysis procedure described by Särndal et al. (1992, section 7.10.2):

$$SE\left(\widehat{O}\right) = \sqrt{\sum_{i=1}^{q} W_i^2 \widehat{U}_i \frac{1-\widehat{U}_i}{n_{i\cdot}-1}}, \tag{23}$$

and

$$SE\left(\widehat{P}_j\right) = \frac{1}{\widehat{N}_{\cdot j}} \sqrt{\frac{N_{j\cdot}^2 \left(1-\widehat{P}_j\right)^2 \widehat{U}_j \left(1-\widehat{U}_j\right)}{n_{j\cdot}-1} + \widehat{P}_j^2 \sum_{i\neq j}^{q} N_{i\cdot}^2 \frac{n_{ij}}{n_{i\cdot}} \cdot \frac{1-\frac{n_{ij}}{n_{i\cdot}}}{n_{i\cdot}-1}} \tag{24}$$

where $n_{i\cdot}$ is the total number of sample units in map class i,

$$\widehat{N}_{\cdot j} = \sum_{i=1}^{q} (N_{i\cdot}/n_{i\cdot}) n_{ij} \tag{25}$$

is the estimated marginal total number of pixels of reference class j, and $N_{j\cdot}$ is the marginal total of map class j. Note that $N_{j\cdot}$ is not an estimate, but is calculated from the full map product. The poststratified estimator for the proportion of area occupied by class k is

$$\widehat{p}_{\cdot k} = \sum_{i=1}^{q} \widehat{p}_{ik} = \sum_{i=1}^{q} W_i \frac{n_{ik}}{n_{i\cdot}} \tag{26}$$

Note that this is the same as the column margin totals for the tables below. The standard error is

$$SE\left(\widehat{p}_{\cdot k}\right) = \sqrt{\sum_{i=1}^{q} W_i^2 \frac{n_{ik}}{n_{i\cdot}} \cdot \frac{1-\frac{n_{ik}}{n_{i\cdot}}}{n_{i\cdot}-1}} = \sqrt{\sum_{i=1}^{q} \frac{W_i \widehat{p}_{ik} - \widehat{p}_{ik}^2}{n_{i\cdot}-1}} \tag{27}$$

References

Särndal, C. E., Swensson, B., & Wretman, J. (1992). *Model-Assisted Survey Sampling*. Springer-Verlag: New York.

Stehman, S.V. (1997). Estimating standard errors of accuracy assessment statistics under cluster sampling. Remote Sensing of Environment, 60, 258-269.

References

Anderson, J.R., Hardy, E.E., Roach, J.T., Witmer, R.E., 1976. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. U.S.G. Survey (Ed.). Government Printing Office, Washington, DC.

Bontemps, S., Herold, M., Kooistra, L., van Groenestijn, A., Hartley, A., Arino, O., Moreau, I., Defourny, P., 2011. Revisiting land cover observations to address the needs of the climate modelling community. Biogeosci. Discuss. 8, 7713.

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.

Brown, J.F., Tollerud, H.J., Barber, C.P., Zhou, Q., Dwyer, J.L., Vogelmann, J.E., Loveland, T.R., Woodcock, C.E., Stehman, S.V., Zhu, Z., Pengra, B.W., Smith, K., Horton, J.A., Xian, G., Auch, R.F., Sohl, T.L., Sayler, K.L., Gallant, A.L., Zelenak, D., Reker, R.R., Rover, J., 2020. Lessons learned implementing an operational continuous United States national land change monitoring capability: The Land Change Monitoring, Assessment, and Projection (LCMAP) approach. Remote Sens. Environ. 238, 111356.

Büttner, G., 2014. CORINE land cover and land cover change products. In: Manakos, I., Braun, M. (Eds.), Land Use and Land Cover Mapping in Europe: Practices & Trends. Springer Netherlands, Dordrecht, pp. 55–74.

Büttner, G., Kosztra, B., Soukup, T., Sousa, A., Langanke, T., 2017. CLC2018 Technical Guidelines. Technical Report. E.E. Agency (Ed.). European Environment Agency, Vienna, Austria.

Calderón-Loor, M., Hadjikakou, M., Bryan, B.A., 2021. High-resolution wall-to-wall land-cover mapping and land change assessment for Australia from 1985 to 2015. Remote Sens. Environ. 252, 112148.

Card, D.H., 1982. Using known map category marginal frequencies to improve estimates of thematic map accuracy. Photogramm. Eng. Remote Sens. 48, 431–439.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, San Francisco, California, USA, pp. 785–794.

Cohen, W.B., Yang, Z., Kennedy, R., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync — Tools for calibration and validation. Remote Sens. Environ. 114, 2911–2924.

Czaplewski, R.L., 2003. Accuracy assessment of maps of forest condition. In: Wulder, M. A., Franklin, S.E. (Eds.), Remote Sensing of Forest Environments. Springer, Boston, MA.

de Gruijter, J.J., ter Braak, C.J.F., 1990. Model-free estimation from spatial samples: a reappraisal of classical sampling theory. Math. Geol. 22, 407–415.

Dwyer, J.L., Roy, D.P., Sauer, B., Jenkerson, C.B., Zhang, H.K., Lymburner, L., 2018. Analysis ready data: enabling analysis of the Landsat archive. Remote Sens. 10, 1363.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global consequences of land use. Science 309, 570–574.

Foody, G.M., 2009. The impact of imperfect ground reference data on the accuracy of land cover change estimation. Int. J. Remote Sens. 30, 3275–3281.

Foody, G.M., 2010. Assessing the accuracy of land cover change with imperfect ground reference data. Remote Sens. Environ. 114, 2271–2285.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850.

Homer, C.G., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G.Z., Coulston, J., Herold, N., 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.

Hu, Y., Hu, Y., 2019. Land cover changes and their driving mechanisms in Central Asia from 2001 to 2017 supported by Google Earth Engine. Remote Sens. 11.
Justice, C., Belward, A., Morisette, J., Lewis, P., Privette, J., Baret, F., 2000. Developments in the 'validation' of satellite sensor products for the study of the land surface. Int. J. Remote Sens. 21, 3383–3390.
Karasiak, N., Dejoux, J.-F., Monteil, C., Sheeren, D., 2021. Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. Mach. Learn. https://doi.org/10.1007/s10994-021-05972-1.
LCMS, 2021. Landscape Change Monitoring System (LCMS). https://www.fs.usda.gov/rmrs/tools/landscape-change-monitoring-system-lcms (last accessed May 28, 2021).
McRoberts, R.E., Stehman, S.V., Liknes, G.C., Næsset, E., Sannier, C., Walters, B.F., 2018. The effects of imperfect reference data on remote sensing-assisted estimators of land cover class proportions. ISPRS J. Photogramm. Remote Sens. 142, 292–300.
Mueller, R., Seffrin, R., 2006. New Methods and Satellites: A Program Update on the NASS Cropland Data Acreage Program. In, Remote Sensing Support to Crop Yield Forecast and Area Estimates. ISPRS, Stresa, Italy.
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.
Omernik, J.M., Griffith, G.E., 2014. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. Environ. Manag. 54, 1249–1266.
Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422.
Pengra, B.W., 2012. One planet, how many people? A review of earth's carrying capacity: a discussion paper for the year of RIO+20. Environ. Dev. 4, 114–135.
Pengra, B.W., Stehman, S.V., Horton, J.A., Dockter, D.J., Schroeder, T.A., Yang, Z., Cohen, W.B., Healey, S.P., Loveland, T.R., 2020a. Quality control and assessment of interpreter consistency of annual land cover reference data in an operational national monitoring program. Remote Sens. Environ. 238, 111261.
Pengra, B.W., Stehman, S.V., Horton, J.A., Dockter, D.J., Schroeder, T.A., Yang, Z., Hernandez, A.J., Healey, S.P., Cohen, W.B., Finco, M.V., Gay, C., Houseman, I.W., 2020b. LCMAP Reference Data Product 1984-2018 Land Cover, Land Use and Change Process Attributes: U.S. Geological Survey Data Release. https://doi.org/10.5066/P9ZWOXJ7 (last accessed 22 July 2021).
Pengra, B.W., Stehman, S.V., Horton, J.A., Wellington, D.F., 2020c. Land Change Monitoring, Assessment, and Projection (LCMAP) Version 1.0 Annual Land Cover and Land Cover Change Validation Tables: U.S. Geological Survey Data Release. https://doi.org/10.5066/P98EC5XR (last accessed 22 July 2021).
Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. 11, 4540.
Pontius Jr., R.G., Krithivasan, R., Sauls, L., Yan, Y., Zhang, Y., 2017. Methods to summarize change among land categories across time intervals. J. Land Use Sci. 12, 218–230.
Radoux, J., Bogaert, P., 2017. Good practices for object-based accuracy assessment. Remote Sens. 9, 646.
Särndal, C.E., Swensson, B., Wretman, J., 1992. Model-Assisted Survey Sampling. Springer-Verlag, New York.
Souza, C.M., Shimbo, J.Z., Rosa, M.R., Parente, L.L., Alencar, A.A., Rudorff, B.F.T., Hasenack, H., Matsumoto, M., Ferreira, L.G., Souza-Filho, P.W.M., de Oliveira, S.W., Rocha, W.F., Fonseca, A.V., Marques, C.B., Diniz, C.G., Costa, D., Monteiro, D., Rosa, E.R., Vélez-Martin, E., Weber, E.J., Lenti, F.E.B., Paternost, F.F., Pareyn, F.G.C., Siqueira, J.V., Viera, J.L., Ferreira Neto, L.C., Saraiva, M.M., Sales, M.H., Salgado, M.P.G., Vasconcelos, R., Galano, S., Mesquita, V.M., Azevedo, T., 2020. Reconstructing three decades of land use and land cover changes in Brazilian biomes with Landsat Archive and Earth Engine. Remote Sens. 12, 2735.
Stehman, S.V., 1992. Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data. Photogramm. Eng. Remote Sens. 58, 1343–1350.
Stehman, S.V., 1997. Estimating standard errors of accuracy assessment statistics under cluster sampling. Remote Sens. Environ. 60, 258–269.
Stehman, S.V., 2000. Practical implications of design-based sampling inference for thematic map accuracy assessment. Remote Sens. Environ. 72, 35–45.
Stehman, S.V., 2001. Statistical rigor and practical utility in thematic map accuracy assessment. Photogramm. Eng. Remote Sens. 67, 727–734.
Stehman, S.V., 2009. Sampling designs for accuracy assessment of land cover. Int. J. Remote Sens. 30, 5243–5272.
Stehman, S.V., 2013. Estimating area from an accuracy assessment error matrix. Remote Sens. Environ. 132, 202–211.
Stehman, S.V., Czaplewski, R.L., 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. Remote Sens. Environ. 64, 331–344.
Stehman, S.V., Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. Remote Sens. Environ. 231, 111199.
Stehman, S.V., Wickham, J.D., 2011. Pixels, blocks of pixels, and polygons: choosing a spatial unit for thematic accuracy assessment. Remote Sens. Environ. 115, 3044–3055.
U.S. Geological Survey, 2019. Joint Response Design for TimeSync Reference Data Collection. U.S Geological Survey, Sioux Falls, SD. www.usgs.gov/media/files/joint-response-design-timesync-reference-data-collection (last accessed July 30, 2021).
U.S. Geological Survey, 2020a. LCMAP Science Product Guide. USGS Land Change Monitoring, Assessment, and Projection (LCMAP) Collection 1 Science Product Guide (SPG). U.S. Geological Survey, Sioux Falls, SD. www.usgs.gov/media/files/lcmap-collection-11-science-product-guide (last accessed July 30, 2021).
U.S. Geological Survey, 2020b. Land Change Monitoring, Assessment, and Projection (LCMAP) Continuous Change Detection and Classification (CCDC) Algorithm Description Document (ADD). U.S. Geological Survey, Sioux Falls, SD. www.usgs.gov/media/files/lcmap-ccdc-add (last accessed July 30, 2021).
U.S. Geological Survey, 2020c. USGS EROS Archive - LCMAP- Continuous Change Detection Classification (CCDC) Products (U.S.), Collection 1.0. U.S. Geological Survey, Sioux Falls, SD. https://doi.org/10.5066/P9W1TO6E.
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.
Wickham, J., Stehman, S.V., Sorenson, D.G., Gass, L., Dewitz, J.A., 2021. Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States. Remote Sens. Environ. 257, 112357.
Ying, Q., Hansen, M.C., Potapov, P.V., Tyukavina, A., Wang, L., Stehman, S.V., Moore, R., Hancher, M., 2017. Global bare ground gain from 2000 to 2012 using Landsat imagery. Remote Sens. Environ. 194, 161–176.
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.