



Thematic accuracy assessment of the 2011 National Land Cover Database (NLCD)



James Wickham^{a,*}, Stephen V. Stehman^b, Leila Gass^c, Jon A. Dewitz^d, Daniel G. Sorenson^e, Brian J. Granneman^f, Richard V. Poss^{g,1}, Lori A. Baer^g

^a Systems Exposure Division, National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27111, USA

^b College of Environmental Science and Forestry, State University of New York, Syracuse, NY 13210, USA

^c U.S. Geological Survey, 520 N. Park Ave., Tucson, AZ 85179, USA

^d Earth Resources and Observation Science (EROS) Center, U.S. Geological Survey, 47914 252nd St., Sioux Falls, SD 57198, USA

^e U.S. Geological Survey, 909st Ave., Seattle, WA 98104, USA

^f SGT, Inc., Contractor to the USGS Earth Resources and Observation Science (EROS) Center, U.S. Geological Survey, 47914 252nd St., Sioux Falls, SD 57198, USA

^g U.S. Geological Survey, Box 25406, Denver Federal Center, MS 980, Denver, CO 80225, USA

ARTICLE INFO

Article history:

Received 31 August 2016

Received in revised form 12 December 2016

Accepted 30 December 2016

Available online 5 February 2017

Keywords:

Forest disturbance

Land-cover change accuracy

MRLC

Stratified sampling

Urbanization

ABSTRACT

Accuracy assessment is a standard protocol of National Land Cover Database (NLCD) mapping. Here we report agreement statistics between map and reference labels for NLCD 2011, which includes land cover for ca. 2001, ca. 2006, and ca. 2011. The two main objectives were assessment of agreement between map and reference labels for the three, single-date NLCD land cover products at Level II and Level I of the classification hierarchy, and agreement for 17 land cover change reporting themes based on Level I classes (e.g., forest loss; forest gain; forest, no change) for three change periods (2001–2006, 2006–2011, and 2001–2011). The single-date overall accuracies were 82%, 83%, and 83% at Level II and 88%, 89%, and 89% at Level I for 2011, 2006, and 2001, respectively. Many class-specific user's accuracies met or exceeded a previously established nominal accuracy benchmark of 85%. Overall accuracies for 2006 and 2001 land cover components of NLCD 2011 were approximately 4% higher (at Level II and Level I) than the overall accuracies for the same components of NLCD 2006. The high Level I overall, user's, and producer's accuracies for the single-date eras in NLCD 2011 did not translate into high class-specific user's and producer's accuracies for many of the 17 change reporting themes. User's accuracies were high for the no change reporting themes, commonly exceeding 85%, but were typically much lower for the reporting themes that represented change. Only forest loss, forest gain, and urban gain had user's accuracies that exceeded 70%. Lower user's accuracies for the other change reporting themes may be attributable to the difficulty in determining the context of grass (e.g., open urban, grassland, agriculture) and between the components of the forest-shrubland-grassland gradient at either the mapping phase, reference label assignment phase, or both. NLCD 2011 user's accuracies for forest loss, forest gain, and urban gain compare favorably with results from other land cover change accuracy assessments.

© 2016 Published by Elsevier Inc.

1. Introduction

The National Land Cover Database (NLCD), sponsored by the MultiResolution Land Characteristics (MRLC) Consortium (<http://www.mrlc.gov>), is a well-established and widely used source of information on land cover (Wickham et al., 2014). The most recent release of the product, NLCD 2011 (Homer et al., 2015), includes 16 land cover classes (http://www.mrlc.gov/nlcd11_leg.php) and related

information for three eras (2001, 2006, 2011) at the native 30 m × 30 m pixel size of Landsat Thematic Mapper. One objective of the NLCD project is to provide land cover monitoring data that can be used to assess land cover change and trends, and the release of NLCD 2011 is the first realization of the database that can be used to assess change over multiple time intervals (Homer et al., 2015).

Accuracy assessment is one of the protocols of the NLCD program. Continuing this protocol of documenting accuracy of NLCD products, the two main objectives of this assessment are: 1) assess the accuracy of the single-date land cover maps produced for each NLCD era (2001, 2006, 2011) at Level II and I classification hierarchies, and 2) assess the accuracy of land cover change across the three NLCD change periods (2001–2006, 2006–2011, 2001–2011). The focus on the accuracy of

* Corresponding author at: National Exposure Research Laboratory, U.S. Environmental Protection Agency, MD: D343-05, RTP, NC 27711, USA.

E-mail address: wickham.james@epa.gov (J. Wickham).

¹ Retired.

change across the three NLCD time periods is consistent with the format used to report NLCD 2006 land cover thematic accuracy (Wickham et al., 2013). NLCD 2006 (Fry et al., 2011) was the first NLCD database to incorporate land cover change. This accuracy assessment was undertaken to document product quality, inform production of future NLCD products, and support monitoring, modeling, and assessments that use NLCD 2011 land cover data.

The continuing development of the NLCD database results in new versions of previously released land cover products. The NLCD 2011 database includes version 1 of the year 2011, version 2 of the year 2006 and version 3 of the year 2001. Thus, the NLCD 2011 accuracy assessment reported in this paper evaluates version 3 of year 2001, version 2 of year 2006 and version 1 of year 2011. Users of NLCD 2001 (Homer et al., 2007) and NLCD 2006 (Fry et al., 2011) products should refer to their associated accuracy assessments when using those products. The accuracy assessment of NLCD 2001, which includes version 1 of NLCD 2001, is reported in Wickham et al. (2010), and the accuracy assessment of NLCD 2006, which includes version 2 of year 2001 and version 1 of year 2006, is reported in Wickham et al. (2013). NLCD 1992 (Vogelmann et al., 2001) is not considered part of the NLCD time series because of substantial methodological differences from later NLCD versions (Homer et al., 2004). The NLCD 1992 accuracy assessments are reported in Stehman et al. (2003) and Wickham et al. (2004).

In addition to the three eras of land cover, the NLCD database also includes percentage urban impervious cover for 2001, 2006, and 2011 (Xian et al., 2011), and forest canopy density for 2001 and 2011 (Coulston et al., 2012; Homer et al., 2007). The number of accuracy assessment objectives increases with the continued growth and development of the NLCD database, and all of these objectives cannot be accommodated with the limited NLCD resources (Stehman et al., 2008). We focus here on accuracy of land cover and land cover change among the three NLCD eras because it was considered the highest priority among MRLC participants. Accuracy of urban impervious cover and forest canopy density are not addressed in this assessment.

2. Methods

2.1. Sampling design

Accuracy assessment methods were based on the sampling design, response design, and analysis components developed by Stehman and Czaplewski (1998). We implemented a stratified random sampling design to accommodate the dual objectives of individual era (i.e., single date) assessments at Level II and Level I (Table 1) and temporal change assessments at Level I for multiple change periods. The continental United States was first divided into east and west regions to create two geographic strata (Fig. 1). This regional stratification was used because previous NLCD accuracy assessments have shown geographic variations in accuracies in which class-specific accuracies tend to be higher when the class was dominant regionally (Stehman et al., 2003; Wickham et al., 2004, 2010, 2013). Thirty-eight (38) strata were sampled within each region, with 16 of these strata corresponding to mapped no change over all three dates for the 16 Level II classes. The other 22 strata were defined based on mapped change over the three dates (Table 2). The 22 change strata prioritized shifts among forest, shrubland, grassland and urban among the 504 possible change combinations of eight Level I classes for three dates (excluding Level I no change classes). The 38 strata accounted for all pixels in the NLCD 2011 map area thereby satisfying one condition of a probability sampling design which is that each pixel in the population must have a non-zero inclusion probability (Stehman, 2001). Accuracy estimates for the temporal component of NLCD 2011 were produced for 17 reporting themes that were based on the eight Level I classes (Table 3). These reporting themes are same as those used in the NLCD 2006 accuracy assessment (Wickham et al., 2013) facilitating comparison of accuracy of NLCD 2011 with NLCD 2006.

Table 1

National Land Cover Database (NLCD) land cover legend for Level II of the classification hierarchy and (class codes). Level I classes are based on the tens digit of the class code, e.g., classes 11 and 12 combine to form class = 10 (water). See http://www.mrlc.gov/nlcd11_leg.php for a complete description of NLCD classes.

Class (code)	Description
Water (11)	Open water, with generally < 25% vegetation or soil cover
Perennial ice/snow (12)	>25% permanent ice or snow
Developed, open space (21)	Dominated by vegetation; impervious cover (IC) ≤ 20%
Developed, low intensity (22)	Mixture of vegetation and IC (20% < IC ≤ 49%)
Developed, medium intensity (23)	Mixture of vegetation and IC (50% < IC ≤ 79%)
Developed, high intensity (24)	Mixture of vegetation and IC (IC ≥ 80%)
Barren (31)	Bedrock, desert pavement, etc.; vegetation <15 cover
Deciduous forest (41)	Trees >20% cover of which >75% shed foliage seasonally
Evergreen forest (42)	Trees >20% cover of which >75% maintain foliage year round
Mixed forest (43)	Trees >20% cover; neither deciduous or evergreen >75% cover
Shrubland (52)	Woody species <5 m and >20% cover
Grassland (71)	Herbaceous cover ≥80%; no management (e.g., tilling) evident
Pasture (81)	Herbaceous cover >20% for livestock, seed, or hay crops
Cultivated crops (82)	Herbaceous or woody cover ≥20% (e.g., corn, orchards)
Woody wetlands (90)	Woody cover >20% on periodically saturated soil
Herbaceous wetland (95)	Herbaceous cover >80% on periodically saturated soil

Previous NLCD accuracy assessments used 10 geographic strata (regions), but only two regions were defined for this assessment because limited resources reduced the total sample size to 8000 from 15,000 sample pixels used in the NLCD 2001 (Wickham et al., 2010) and NLCD 2006 (Wickham et al., 2013) accuracy assessments. The eastern U.S. region received 3900 sample pixels and the western U.S. region received 4100 sample pixels. There were no sample pixels of the NLCD perennial ice and snow class in the eastern region.

2.2. Response design

The main elements of the response design were: 1) blind interpretation; 2) reliance on Google Earth™ time series imagery to determine the reference labels; 3) reliance on the pixel as the spatial support unit of the assessment (Stehman and Wickham, 2011); 4) assignment of primary and alternate reference labels, and; 5) specific rules for coding

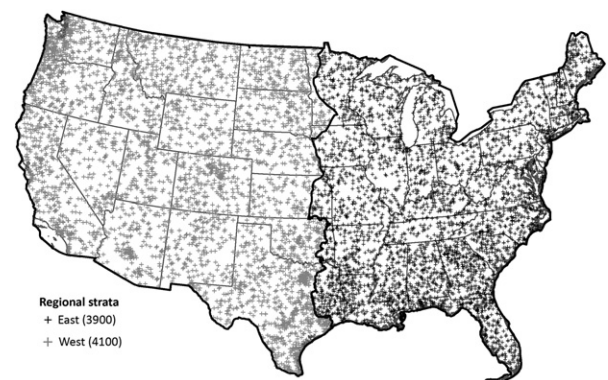


Fig. 1. NLCD 2011 accuracy assessment sample pixel locations and regional strata. The east-west regional strata were based on the mapping regions developed for NLCD 2001, version 1 (Homer and Gallant, 2001).

Table 2

Sample strata within each of the two geographic strata. Strata 1 through 16 are class-specific “no change” strata based on the Level II classes and strata 17 through 37 are change strata based on the Level I classes. The “catchall” stratum includes all other three-date Level I land cover class combinations. Numbers in parentheses are total sample size across both geographic strata.

Strata (2001–2006–2011)	Strata (continued)
1) Water–water–water (165)	20) Forest–forest–urban (150)
2) Ice–ice–ice (25)	21) Forest–grassland–grassland (80)
3) Open urban (OU)–OU–OU (260)	22) Forest–forest–grassland (180)
4) Low density urban (LDU)–LDU–LDU (200)	23) Shrubland–forest–forest (80)
5) Medium density urban (MDU)–MDU–MDU (180)	24) Shrubland–shrubland–forest (165)
6) High density urban (HDU)–HDU–HDU (165)	25) Shrubland–grassland–grassland (80)
7) Barren–barren–barren (225)	26) Shrubland–shrubland–grassland (165)
8) Deciduous forest (DF)–DF–DF (550)	27) Shrubland–urban–urban (80)
9) Evergreen forest (EF)–EF–EF (565)	28) Shrubland–shrubland–urban (150)
10) Mixed forest (MF)–MF–MF (220)	29) Grassland–shrubland–shrubland (80)
11) Shrubland–shrubland–shrubland (615)	30) Grassland–grassland–shrubland (165)
12) Grassland–grassland–grassland (490)	31) Grassland–urban–urban (80)
13) Pasture–pasture–pasture (475)	32) Grassland–grassland–urban (150)
14) Crop–crop–crop (660)	33) Agriculture–urban–urban (80)
15) Woody wetland (WW)–WW–WW (265)	34) Agriculture–agriculture–urban (150)
16) Emergent wetland (EM)–EM–EM (180)	35) Forest–shrubland–grassland (80)
17) Forest–shrubland–shrubland (80)	36) Grassland–grassland–agriculture (150)
18) Forest–forest–shrubland (180)	37) Grassland–forest–forest (80)
19) Forest–urban–urban (80)	38) Catchall (275)

primary and alternate reference labels across Level II and Level I classification hierarchies. Collection of reference labels was accomplished by four persons at the U.S. Geological Survey. Before assigning reference labels to the actual sample pixels, interpreters completed training and orientation to promote consistency among interpreters and gain experience in collection of reference labels for some of the common land cover trends in the NLCD maps (Mann and Rothley, 2006). Landsat path/rows in the vicinity of Jacksonville, Florida and Denver, Colorado were used for training and orientation. Following training and orientation, reference label collection was initiated with 200 sample pixels that were interpreted collectively by all four interpreters to further enhance consistency among interpreters (Mann and Rothley, 2006), and following completion of the interpretation of these sample pixels, each person was assigned an additional 1950 sample pixels that they interpreted individually. Weekly web-enabled conference calls were conducted during the collection of reference labels to further ensure consistent interpretation.

Reference labels were collected by the interpreters without knowledge of the map classification (response design element 1). Each interpreter was provided three vector Keyhole Markup language Zipped (KMZ) files of the sample pixels for overlay on Google Earth™ imagery. The vector files were point and polygon expressions of the sample pixels, and a vector file of the 3×3 pixel window surrounding the sample pixel. The 3×3 pixel window file was supplied to add context; it is appropriate to survey the surrounding landscape to determine the most appropriate labels for a sample pixel (Stehman and Czaplewski, 1998). The vector files were overlaid on the Google Earth™ time series imagery to assist the interpreters in obtaining the reference label for the sample pixel (response design element 2). The interpreters also had Landsat imagery acquisition dates for the NLCD classifications to guide selection of the most appropriate Google Earth™ date to use when determining the reference label. The goal of reference label assignment was to identify the most appropriate land cover labels that corresponded to the ground condition for the sample pixel (Stehman and Wickham, 2011) (response design element 3).

The interpreters collected primary and alternate reference labels at Level II and Level I of the NLCD classification hierarchy for each sample pixel while keeping in mind the NLCD mapping protocols. The primary

Table 3

Reporting themes for accuracy results.

Reporting themes	Description
1) Water loss	From water to any other class (2001–2006, 2006–2011, 2001–2011)
2) Water gain	To water from any other class (2001–2006, 2006–2011, 2001–2011)
3) Urban gain	To urban from any other class (2001–2006, 2006–2011, 2001–2011)
4) Forest loss	From forest to any other class (2001–2006, 2006–2011, 2001–2011)
5) Forest gain	To forest from any other class (2001–2006, 2006–2011, 2001–2011)
6) Shrubland loss	From shrubland to any other class (2001–2006, 2006–2011, 2001–2011)
7) Shrubland gain	To shrubland from any other class (2001–2006, 2006–2011, 2001–2011)
8) Grassland loss	From grassland to any other class (2001–2006, 2006–2011, 2001–2011)
9) Grassland gain	To grassland from any other class (2001–2006, 2006–2011, 2001–2011)
10) Agriculture loss	From agriculture to any other class (2001–2006, 2006–2011, 2001–2011)
11) Agriculture gain	To agriculture from any other class (2001–2006, 2006–2011, 2001–2011)
12) Water–no change	Water across all three NLCD eras
13) Urban–no change	Urban across all three NLCD eras
14) Forest–no change	Forest across all three NLCD eras
15) Shrubland–no change	Shrubland across all three NLCD eras
16) Grassland–no change	Grassland across all three NLCD eras
17) Agriculture–no change	Agriculture across all three NLCD eras

label was that deemed most correct and the alternate label was considered a very likely alternative (response design element 4). An alternate label was not assigned if, in the interpreter's judgment, the primary class was the only possible class. In aggregate for the three dates sampled, no alternate label was assigned for 42% of the sample pixels at Level II and 65% of the sample pixels at Level I. Use of primary and alternate labels was consistent with all previous NLCD accuracy assessments (Stehman et al., 2003; Wickham et al., 2004, 2010, 2013), and can be considered a special case of the linguistic scale, fuzzy membership analysis (Stehman et al., 2003, p. 513) reported in Gopal and Woodcock (1994). The main protocol for collection of reference data was for each interpreter to examine the time series of Google Earth™ imagery and determine the primary and alternate reference label sets at Level I for all three eras. The interpreters then used the Level I reference labels to assign the Level II reference labels (i.e., the Level II label had to be one of the subclasses within the Level I hierarchy).

Reference labels were assigned using the conceptual model of NLCD mapping protocols (response design element 5), rather than from the perspective of the land cover evident on Google Earth™ imagery (Comber et al., 2005). The numerous forest fires that have occurred in the western United States over the past decade provide a good example of the difference between reference label assignments from the perspective of NLCD mapping protocols versus the perspective of the land cover evident on Google Earth™ imagery. Many of these areas impacted by forest fire are comprised of standing dead trees, and thus from the Google Earth™ perspective there would be a tendency to label sample pixels in such areas as forest since trees are still present and ecological succession is likely to follow. The NLCD protocol was to map areas that changed from forest to burned forest as forest to shrubland so the reference label assignment protocol implemented would label such a case as forest in 2006 and shrubland in 2011. Reference label assignment accounted for such protocols and was conducted by interpreters who also participated in production of the NLCD maps.

2.3. Analysis

The analysis component employed general estimation theory of probability sampling (cf. Särndal et al., 1992). The sample-based

estimates incorporate the known inclusion probabilities of the stratified random design (Stehman, 2001; Stehman and Czaplewski, 1998) although special case estimation formulas are used that do not show the inclusion probabilities explicitly. Overall accuracy was estimated as

$$\hat{O} = \left(\frac{1}{N} \right) \sum_{h=1}^H N_h \hat{p}_h \quad (1)$$

where \hat{p}_h is the sample proportion of pixels correctly classified in stratum h , N is the total number of pixels in the region, N_h is the population size of stratum h , and the summation is over all H strata ($H = 38$ for a regional estimate and $H = 76$ for a national estimate). Overall accuracy was estimated for the individual, single-date land cover products (2001, 2006, 2011) and the change between them at Level I for the three time intervals (2001–2006, 2006–2011, 2001–2011). User's and producer's accuracies were estimated as a ratio $R = Y/X$, where Y is the population total of y_u where,

$$y_u = \begin{cases} 1 & \text{if pixel } u \text{ satisfies condition A} \\ 0 & \text{if pixel } u \text{ does not satisfy condition A} \end{cases} \quad (2)$$

and X is the population total of x_u , where

$$x_u = \begin{cases} 1 & \text{if pixel } u \text{ satisfies condition B} \\ 0 & \text{if pixel } u \text{ does not satisfy condition B} \end{cases} \quad (3)$$

For example, to estimate user's accuracy for the Level I class forest (e.g., Tables 1, 2), condition A would be that the map and reference labels were both forest, and condition B would be that the map label was forest. The ratio Y/X would then be the parameter defining user's accuracy, which is the total number of pixels in the region for which both the map and reference labels were forest divided by the number of pixels in the region mapped as forest. To estimate producer's accuracy of forest, condition A would remain the same, but condition B would be that the reference label was forest. The combined ratio estimator (Cochran, 1977, Section 6.11) for user's or producer's accuracy is then

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} = \frac{\sum_{h=1}^H N_h \bar{y}_h}{\sum_{h=1}^H N_h \bar{x}_h} \quad (4)$$

where \bar{x}_h is the sample mean of x_u in stratum h (i.e., Table 2) and \bar{y}_h is the sample mean of y_u in stratum h . We report accuracy estimates for agreement based on the map label matching the primary reference label and also for agreement based on the map matching the primary reference label or an alternate reference label. For assessments of change accuracy, as many as three alternate reference conditions were possible. For example, when assessing the 2001 to 2006 NLCD change, the alternate reference labels included the alternate 2001 Level 1 class with the 2006 alternate Level 1 class, the primary 2001 Level I class with the alternate 2006 Level I class, and the alternate 2001 class with the primary 2006 class. These three comparisons were in addition to the comparison using the primary 2001 Level I class and the primary 2006 Level I class to determine the reference class of change.

The estimated variance of the combined ratio estimator is

$$\hat{V}(\hat{R}) = \left(\frac{1}{\hat{X}^2} \right) \left[\sum_{h=1}^H N_h^2 \left(1 - n_h/N_h \right) \left(s_{y_h}^2 + \hat{R}^2 s_{x_h}^2 - 2\hat{R}s_{xyh} \right) / n_h \right] \quad (5)$$

where n_h is the sample size in stratum h , $s_{y_h}^2$ and $s_{x_h}^2$ are the sample variances of y_u and x_u for stratum h and s_{xyh} is the sample covariance for y_u and x_u for stratum h . Sample data from several strata may contribute to the accuracy estimators for a targeted class (Table 2) because the strata do not always directly correspond to a target class. Estimation of user's accuracy for shrubland loss during 2001 to 2006, for example, would include sample pixels from strata 23 through 28 in Table 2. The values of y_u , \bar{y}_h , and $s_{y_h}^2$ equal zero (0) for a stratum in which no sample pixels

satisfy condition A (the condition defining the numerator of \hat{R}), and, similarly, the values of x_u , \bar{x}_h , and $s_{x_h}^2$ equal zero (0) for a stratum in which no pixels satisfy condition B (the condition defining the denominator of \hat{R}). Estimates were computed using version 9.3 of SAS (Statistical Analysis Software, SAS, Inc., Cary, North Carolina, USA).

We used a nominal benchmark of 85% as a quality threshold for interpreting agreement between map and reference data (Anderson et al., 1976). We recognize that this benchmark has been used uncritically as a heuristic, and its use may not be appropriate in all contexts (Foody, 2006). Nevertheless, we feel that it serves as a useful guide for evaluation of the quality of the temporal NLCD maps.

3. Results

3.1. Accuracy of single-date maps

Unless otherwise stated, the results presented are based on the definition of agreement as a match between the map label and either the primary or alternate reference label. At Level II of the classification hierarchy, land cover overall accuracies of the NLCD 2011 individual date products were 82% for 2011 (Table 4) and 83% for both 2006 and 2001 (Tables 5, 6). High user's accuracies ($\geq 85\%$) were realized for water (11), high intensity developed (24), deciduous forest (41), evergreen forest (42), shrubland (52), and cropland (82) when agreement was defined as a match between the map and the primary or alternate reference label. There was a regional dichotomy in Level II overall accuracy. Level II overall accuracies for 2011 were approximately 10% higher in the western sampling region than the eastern sampling region, primarily from much higher agreement in the western region for shrubland and grassland as well as the urban classes (Tables 7, 8). A similar east versus west difference in overall accuracy was observed for 2001 and 2006 (tables not included).

Overall accuracies increased from 6% to 9% across all NLCD eras when land cover classes were aggregated from Level II to Level I, depending on the definition of agreement (Tables 9–11). Level I overall accuracies were about 9% higher than the Level II overall accuracies when the definition of agreement was restricted to a match between the map label and the primary reference label only. High user's accuracies ($\geq 85\%$) were realized for water (10), forest (40), shrubland (50), and agriculture (80) across all NLCD eras. Overall accuracy was approximately 6% higher in the east than in the west when agreement was defined as a match between the map label and primary reference label only (Tables 12, 13).

Map homogeneity and the definition of agreement had substantial impacts on overall accuracy. Constraining agreement to a match between the map and primary reference label reduced overall accuracies from 9% to 15% relative to overall accuracy based on agreement defined as a match based on either the primary or alternate reference label (Table 14). The magnitude of the change in accuracy depended on the NLCD era, level of classification hierarchy, and sampling region. The impact of map homogeneity, defined here as like-classified pixels (Level I) for a sample pixel's eight immediate neighbors, was similar to the impact of agreement definition. Depending on the NLCD era, level of classification hierarchy, sampling regions, and agreement definition, overall accuracy improved by 4–13% when only the subset of sample pixels with like-classified neighbors was considered.

3.2. Accuracy of change

Overall accuracies for a binary change versus no change classification exceeded 95% for all three change periods (Tables 15–17). User's and producer's accuracies for no change were $>95\%$ in all cases, but accuracy of change was lower. User's accuracy of change was approximately 55% for all change periods when agreement was defined as a match with only the primary reference change labels, and increased to

Table 4

Agreement between map and reference labels for NLCD 2011 for the continental United States at Level II of the classification hierarchy. Agreement was defined as a match between the primary and map reference labels. Cell entries represent percent of area, and 0.0000 denotes a non-zero value <0.00005 . Sample size is reported in the column and row labeled n. Producer's Accuracy (Prod), User's accuracy (User) and the standard errors (in parentheses) are rounded to the nearest whole number. The labels Auser and Aprod are the User's and Producer's accuracies with agreement defined as a match between the map and either the primary or alternate reference labels. OA_1 is overall accuracy for agreement defined as a match between the map and primary reference labels, and OA_2 is overall accuracy for agreement defined as a match between the map and either the primary or alternate reference labels. $OA_1 = 65.8 (\pm 0.7\%)$ and $OA_2 = 82\% (\pm 0.5\%)$.

Map ↓	Reference																	Total	User	Auser	n
	11	12	21	22	23	24	31	41	42	43	52	71	81	82	90	95					
11	1.5683	0.0037	0.0118	0.0197		0.0118	0.0041	0.0164			0.0079		0.0086	0.0125	0.0517	0.0486	1.7573	89 (2)	92 (2)	189	
12			0.0185	20 (8)	36 (10)	25															
21	0.0003		1.2173	0.5815	0.0839	0.0006		0.2898	0.1927	0.0162	0.1259	0.1612	0.3301	0.3187	0.0306	0.0000	3.3486	36 (3)	57 (3)	593	
22	0.0042		0.4068	0.6818	0.3293	0.0057	0.0147	0.0162	0.0003	0.0059	0.0144	0.0261	0.0266		0.0106	1.5379	44 (4)	69 (3)	517		
23	0.0005		0.0422	0.1259	0.3335	0.1522	0.0095	0.0039	0.0041		0.0002		0.0004			0.6743	50 (3)	79 (3)	403		
24	0.0015		0.0158	0.0074	0.0264	0.1935	0.0062	0.0000				0.0000	0.0026			0.2536	76 (3)	83 (3)	245		
31	0.0531		0.0235	0.0039			0.5486	0.0101	0.0132		0.2476	0.2876	0.0047	0.0072	0.0186	0.0101	1.2282	45 (4)	60 (4)	244	
41	0.0234		0.2823	0.0662			0.0098	8.5249	0.7939	0.6794	0.3239	0.0666	0.0467	0.1827	0.2065	0.0336	11.2397	76 (2)	84 (2)	615	
42	0.0129		0.1248	0.0265				0.4159	9.1446	0.4223	1.4967	0.2346	0.0258	0.0144	0.0818	0.0005	12.0008	76 (2)	88 (1)	862	
43			0.0344	0.0049				0.5624	0.6890	0.5980	0.0840	0.0230	0.0115		0.0786		2.0857	29 (3)	59 (3)	235	
52	0.0501		0.3275	0.0501			0.1182	0.6593	1.2929	0.0759	15.4971	3.1980	0.2947	0.0848	0.0184	0.0237	22.1405	69 (2)	88 (1)	1224	
71	0.0388		0.2628	0.0872	0.0341		0.1034	0.3007	0.3088	0.0471	3.5890	7.9595	1.7034	0.6057	0.0198	0.0586	15.1190	53 (2)	81 (1)	1022	
81	0.0168		0.4083	0.0774			0.0168	0.4668	0.0813	0.0168	0.1587	0.3044	3.8932	1.3442	0.0335	0.0894	6.9075	56 (2)	72 (2)	514	
82	0.0518		0.4177	0.0477	0.0238	0.0244	0.0005	0.3402	0.0477	0.0238	0.1540	0.2983	1.6167	12.9050	0.0569	0.1104	16.1189	80 (2)	88 (1)	823	
90	0.0748		0.0334	0.0202	0.0202			0.7819	0.4925	0.0404	0.1654	0.0686	0.0132	0.0328	2.1472	0.1128	4.0035	54 (3)	70 (3)	283	
95	0.0436		0.0172				0.0044	0.0573	0.0132		0.1000	0.0725	0.0615	0.0131	0.2576	0.6560	1.2963	51 (4)	60 (4)	206	
Total	1.9401	0.0037	3.6258	1.8004	0.8532	0.3881	0.8457	12.4457	13.0741	1.9198	21.9563	13.4139	8.0301	15.5480	3.0010	1.1542	100.0000				
Prod	81 (4)	100 (0)	34 (3)	38 (3)	39 (4)	50 (5)	65 (7)	68 (1)	70 (1)	31 (3)	71 (1)	59 (2)	48 (2)	83 (1)	72 (3)	57 (5)					
Aprod	84 (3)	100 (0)	60 (3)	56 (4)	65 (5)	72 (5)	81 (6)	81 (1)	79 (1)	65 (4)	89 (1)	87 (1)	68 (2)	88 (1)	86 (2)	71 (4)					
n	227	5	601	393	345	284	143	820	1130	158	1198	857	585	876	221	157				8000	

Table 5

Agreement between map and reference labels for NLCD 2006 for the continental United States at Level II of the classification hierarchy. See Table 4 for explanation of contents. $OA_1 = 66.5 (\pm 0.7\%)$ and $OA_2 = 82.8\% (\pm 0.5\%)$.

Map ↓	Reference																				
	11	12	21	22	23	24	31	41	42	43	52	71	81	82	90	95	Total	User	Auser	n	
11	1.5277	0.0037	0.0118	0.0079		0.0118	0.0120	0.0235			0.0079	0.0082	0.0125	0.0079	0.0471	0.0448	1.7231	89 (3)	92 (2)	181	
12								0.0096				0.0052						0.0185	20 (8)	36 (10)	25
21	0.0041			1.2006	0.5789	0.0895	0.0003	0.0006	0.2796	0.1251	0.0155	0.1949	0.1699	0.3384	0.3278	0.0305		3.3556	36 (3)	57 (3)	388
22				0.4360	0.6462	0.3121	0.0023	0.0145	0.0157	0.0002		0.0053	0.0140	0.0213	0.0257		0.0102	1.5034	43 (4)	69 (3)	317
23	0.0005		0.0403	0.1369	0.3012	0.1462	0.0086	0.0040	0.0011		0.0002						0.6289	48 (4)	79 (3)	263	
24	0.0015		0.0122	0.0067	0.0248	0.1806	0.0060						0.0026				0.2343	77 (3)	83 (3)	188	
31	0.0456		0.0148	0.0039			0.5558	0.0147	0.0163		0.2492	0.2863	0.0047	0.0031	0.0179	0.0109	1.2231	45 (4)	61 (4)	243	
41	0.0280		0.2183	0.0678	0.0003	0.0000	0.0100	8.5900	0.7357	0.7079	0.4289	0.0696	0.0757	0.1855	0.2065	0.0361	11.3603	76 (2)	85 (2)	730	
42	0.0159		0.1429	0.0003	0.0001		0.0046	0.4298	9.3704	0.4211	1.4512	0.2828	0.0258	0.0129	0.0798	0.0005	12.2390	77 (2)	88 (1)	1026	
43			0.0231	0.0050				0.5736	0.7184	0.6112	0.0962	0.0264	0.0115		0.0930		2.1584	28 (3)	59 (3)	271	
52	0.0521		0.3374	0.0523	0.0005	0.0003	0.1231	0.6006	1.0155	0.0594	15.8257	3.7683	0.2582	0.0981	0.0204	0.0237	22.2354	71 (2)	89 (1)	1305	
71			0.2733	0.0881	0.0345	0.0002	0.1024	0.3060	0.2505	0.0454	3.5237	7.9991	1.6719	0.6071	0.0127	0.0566	14.9714	53 (2)	82 (2)	1231	
81	0.0168		0.4119	0.0623	0.0009	0.0004		0.4509	0.0813		0.1710	0.3233	3.9544	1.3265	0.0335	0.0894	6.9237	57 (2)	72 (2)	551	
82	0.0766		0.3977	0.0499	0.0247	0.0251	0.0046	0.3349	0.0238	0.0004	0.1340	0.3160	1.6868	12.8945	0.0569	0.0980	16.1238	80 (2)	88 (1)	792	
90	0.0748		0.0330	0.0248	0.0202			0.8105	0.4392	0.0404	0.1890	0.0888	0.0132	0.0328	2.1682	0.1128	4.0477	54 (3)	70 (3)	294	
95	0.0401		0.0132				0.0044	0.0573	0.0132		0.0747	0.0740	0.0574	0.0131	0.2457	0.6606	1.2537	53 (4)	63 (4)	195	
Total	1.8835	0.0037	3.5664	1.7209	0.8088	0.3760	0.8561	12.4911	12.7906	1.9023	22.3518	13.4319	8.1353	15.5350	3.0122	1.1436	100.0000				
Prod	81 (3)	100 (0)	34 (3)	38 (3)	37 (4)	49 (5)	65 (7)	69 (1)	73 (1)	32 (3)	71 (1)	60 (2)	49 (2)	83 (1)	72 (3)	58 (5)					
Aprod	86 (3)	100 (0)	61 (3)	56 (4)	64 (5)	72 (8)	82 (6)	81 (1)	83 (1)	68 (4)	89 (1)	87 (1)	69 (2)	88 (1)	87 (2)	73 (4)					
n	215	5	601	322	251	214	140	874	1184	169	1257	859	651	874	223	161				8000	

Table 6Agreement between map and reference labels for NLCD 2001 for the continental United States at Level II of the classification hierarchy. See Table 4 for explanation of contents. $OA_1 = 67.0 (\pm 0.7\%)$ and $OA_2 = 83.2\% (\pm 0.5\%)$.

Map ↓	Reference																Total	User	Auser	n
	11	12	21	22	23	24	31	41	42	43	52	71	81	82	90	95				
11	1.5695		0.0235	0.0079			0.0118	0.0281			0.0079	0.0079	0.0079		0.0629	0.0368	1.7644	89 (2)	93 (2)	191
12		0.0037					0.0096					0.0052					0.0185	20 (8)	36 (10)	25
21	0.0040		1.1770	0.5573	0.0869			0.2912	0.1533	0.0152	0.1543	0.1645	0.3696	0.3353	0.0304		3.3389	35 (3)	56 (3)	280
22			0.4356	0.6149	0.2836	0.0106	0.0143	0.0243			0.0086	0.0171	0.0227	0.0186		0.0100	1.4602	42 (4)	68 (4)	210
23			0.0344	0.1126	0.2687	0.1372	0.0073	0.0039					0.0039				0.5680	47 (4)	80 (3)	180
24	0.0015		0.0060	0.0076	0.0230	0.1593	0.0060										0.2032	78 (3)	85 (3)	165
31	0.0431		0.0248	0.0039			0.5334	0.0055	0.0117	0.0008	0.2455	0.2791	0.0047	0.0031	0.0179	0.0109	1.1843	45 (4)	62 (4)	234
41	0.0280		0.1950	0.0704	0.0012	0.0009	0.0098	8.7266	0.7001	0.7314	0.3748	0.1476	0.0778	0.1784	0.2086	0.0357	11.4860	76 (2)	86 (1)	780
42	0.0159		0.1117	0.0005	0.0000			0.4178	9.7589	0.4258	1.4730	0.1347	0.0257	0.0149	0.0818		12.4606	78 (2)	89 (1)	1127
43			0.0230	0.0051				0.6063	0.7658	0.5838	0.1284	0.0051	0.0115		0.0930		2.2221	26 (3)	59 (3)	305
52	0.0480		0.3117	0.0522	0.0005	0.0003	0.0784	0.6050	0.9260	0.0691	15.8047	3.7531	0.2577	0.0851	0.0219	0.0322	22.0459	72 (2)	90 (1)	1365
71	0.0002		0.2513	0.0547	0.0002	0.0003	0.1075	0.2915	0.2615	0.0458	3.4782	8.0118	1.6267	0.6366	0.0122	0.0537	14.8322	54 (2)	82 (2)	1228
81	0.0168		0.4025	0.0619	0.0018			0.4571	0.1073		0.1837	0.3610	3.9603	1.3570	0.0335	0.0853	7.0281	56 (2)	72 (2)	604
82	0.0477		0.3920	0.0502	0.0267	0.0004	0.0046	0.3362	0.0238	0.0004	0.0785	0.2731	1.7563	12.9468	0.0477	0.1056	16.0899	81 (2)	89 (1)	818
90	0.0665		0.0336	0.0248	0.0202			0.8301	0.4708	0.0450	0.1302	0.0727	0.0132	0.0384	2.1816	0.1002	4.0281	54 (3)	71 (3)	288
95	0.0435		0.0132				0.0044	0.0573	0.0132		0.0425	0.0876	0.0699	0.0044	0.2490	0.6855	1.2706	54 (4)	64 (4)	200
Total	1.8845	0.0037	3.4352	1.6240	0.7128	0.3089	0.7870	12.6806	13.1924	1.9172	22.1104	13.3206	8.2078	15.6185	3.0404	1.1558	100.0000			
Prod	83 (3)	100 (0)	34 (3)	38 (3)	38 (4)	52 (4)	68 (7)	69 (1)	74 (1)	31 (3)	72 (1)	60 (2)	48 (2)	83 (1)	72 (3)	59 (5)				
Aprod	87 (3)	100 (0)	57 (4)	67 (4)	67 (5)	80 (4)	82 (1)	82 (1)	83 (1)	68 (4)	89 (1)	87 (1)	69 (2)	88 (1)	87 (2)	73 (4)				
n	217	5	444	256	181	179	127	965	1352	190	1203	857	708	930	227	159				8000

Table 7Agreement between map and reference labels for NLCD 2011 for the eastern United States at Level II of the classification hierarchy. See Table 4 for explanation of contents. $OA_1 = 63.0 (\pm 0.9\%)$ and $OA_2 = 76.2\% (\pm 0.8\%)$.

Map ↓	Reference																Total	User	Auser	n
	11	12	21	22	23	24	31	41	42	43	52	71	81	82	90	95				
11	2.3453		0.0291	0.0291		0.0291		0.0405					0.0114	0.0114	0.1278	0.1101	2.7339	86 (3)	89 (3)	100
12																				
21	0.0004		1.6773	1.0680	0.1569			0.6172	0.3791	0.0400	0.0395	0.0083	0.5698	0.5677	0.0757	0.0001	5.2000	32 (4)	55 (4)	318
22	0.0098		0.6978	1.2671	0.5285	0.0092	0.0254	0.0281	0.0002		0.0011	0.0007	0.0527	0.0330		0.0261	2.6797	47 (5)	70 (4)	261
23	0.0010		0.0783	0.1913	0.4359	0.2656	0.0040	0.0096	0.0098		0.0004			0.0009			0.9969	44 (5)	76 (4)	181
24	0.0037		0.0316	0.0112	0.0336	0.3093	0.0154	0.0001				0.0001					0.4050	76 (4)	81 (4)	131
31	0.0479		0.0268	0.0097			0.1016	0.0249	0.0116		0.0210	0.0365	0.0116	0.0077	0.0039	0.0039	0.3070	33 (7)	43 (7)	110
41	0.0578		0.6987	0.1156				19.3560	1.5040	1.5607	0.3168	0.1156	0.1156	0.4274	0.4624	0.0578	24.7884	78 (2)	87 (2)	469
42	0.0318		0.3086					0.8952	6.0005	0.9794	0.3382	0.0993	0.0637	0.0356	0.2023		8.9547	67 (3)	84 (2)	417
43			0.0852					1.1749	1.2954	1.4304	0.0751	0.0568	0.0284		0.1704		4.3165	33 (4)	64 (4)	159
52	0.0074		0.3247	0.0276				0.7548	1.2341	0.1849	0.6731	0.3189	0.2216	0.0755	0.0454	0.0581	3.9260	17 (2)	28 (3)	399
71			0.2273	0.0483	0.0006		0.0024	0.4884	0.3147	0.0305	0.3525	0.5944	0.7097	0.2939	0.0469	0.0311	3.1408	19 (2)	39 (4)	346
81	0.0415		0.8822	0.1659			0.0415	1.0784	0.1244	0.0415	0.1776	0.1358	7.3222	2.3354	0.0829	0.0829	12.5122	59 (3)	75 (3)	325
82	0.1180		0.4834	0.1180	0.0590			0.7199	0.1180	0.0590	0.1885	0.0595	1.6042	15.3558	0.1408	0.0704	19.0945	80 (2)	86 (2)	396
90	0.1098		0.0500	0.0500	0.0500			1.7608	1.1224	0.1000	0.2613	0.0500		0.0500	5.0581	0.1711	8.8334	57 (4)	74 (3)	183
95	0.0549		0.0316					1.088	0.0218		0.1496	0.0294	0.0653	0.0114	0.5279	1.1102	2.1109	53 (5)	61 (5)	105
Total	2.8294		5.6326	3.1019	1.2646	0.6132	0.1903	27.0575	12.1360	4.4263	2.5950	1.5052	10.7763	19.2056	6.9445	1.7218	100.0000			
Prod	83 (4)		30 (3)	41 (4)	35 (5)	50 (5)	53 (15)	72 (2)	49 (2)	32 (4)	26 (4)	40 (6)	68 (3)	80 (2)	73 (3)	64 (6)				
Aprod	87 (4)		54 (4)	59 (5)	61 (6)	75 (7)	60 (14)	82 (1)	62 (2)	65 (4)	48 (5)	65 (6)	79 (2)	86 (2)	87 (2)	76 (6)				
n	115		355	227	142	137	50	649	641	144	234	165	327	457	179	78				3900

Table 8
Agreement between map and reference labels for NLCD 2011 for the western United States at Level II of the classification hierarchy. See Table 4 for explanation of contents. OA₁ = 67.8 (±1.0%); OA₂ = 86.0% (±0.7%).

Map ↓	Reference	11	12	21	22	23	24	31	41	42	43	52	71	81	82	90	95	Total	User	Auser	n
11	1.0413				0.0133			0.0068				0.0133			0.0133		0.0068	1.0950	95 (2)	86 (2)	89
12	0.0062							0.0161				0.0133						0.0310	20 (8)	36 (10)	25
21	0.0003	0.9054		0.9054	0.2515	0.0343	0.0009		0.0678	0.0662		0.1845	0.0087	0.1674	0.1499			2.0931	43 (4)	61 (4)	275
22	0.0004	0.2095		0.2095	0.2849	0.1942	0.0033	0.0074	0.0081	0.0003		0.0091	0.2649	0.0004	0.0223			0.7636	37 (5)	67 (4)	256
23	0.0002	0.0177		0.0177	0.0816	0.2674	0.0754	0.0132	0.0081	0.0002			0.0237	0.0004				0.4555	59 (5)	84 (3)	222
24		0.0051		0.0051	0.0048	0.0216	0.1150							0.0044				0.1509	76 (5)	87 (4)	114
31	0.0567	0.0212		0.0212				0.8517		0.0143		0.4012	0.4579		0.0068	0.0287	0.0143	1.8528	46 (4)	62 (4)	134
41					0.0327			0.0164	1.1801	0.3123	0.0819	0.3287	0.0334		0.0167	0.0329	0.0172	2.0521	58 (4)	68 (4)	146
42					0.0445				0.0909	11.2767	0.4445	2.2824	0.3264				0.0008	14.0644	80 (2)	89 (1)	445
43					0.0082				0.1471	0.2778	0.0334	0.0889	0.0002		0.0163			0.5728	6 (3)	33 (6)	76
52	0.0791	0.3294		0.3294	0.0654			0.1984	0.5945	1.3327	0.0020	25.5495	6.3586	0.3442	0.0911	0.0015	0.0004	34.9452	73 (2)	93 (1)	825
71	0.0651	0.2869		0.2869	0.1136	0.0568		0.1720	0.1734	0.3048	0.0583	5.7838	12.9538	2.3772	0.8172		0.0773	23.2416	56 (2)	85 (2)	676
81		0.0869		0.0869	0.0174				0.0521	0.0521		0.1459	0.4188	1.5679	0.6720		0.0937	3.1608	50 (2)	65 (4)	189
82	0.0068	0.3732		0.3732			0.0410	0.0009	0.0828			0.1306	0.4602	1.6251	11.2430		0.1374	14.1010	80 (2)	89 (2)	427
90	0.0512	0.0222		0.0222					0.1182	0.0654		0.1004	0.0812	0.0222	0.0211	0.1732	0.0733	0.7282	24 (4)	37 (5)	100
95	0.0360	0.0074		0.0074				0.0074	0.0233	0.0074		0.0633	0.1017	0.0589	0.0143	0.0743	0.3480	0.7439	47 (5)	58 (5)	101
Total	1.3370	0.0062		0.0062			0.2355	1.2902	2.5371	13.7103	0.2201	35.0856	21.4893	6.1679	13.0677	0.3268	0.7693	100.0000			
Prod	78 (6)	100 (0)		40 (5)	31 (5)	47 (6)	49 (10)	66 (8)	47 (5)	82 (2)	15 (8)	73 (1)	60 (2)	25 (3)	86 (2)	53 (8)	45 (7)				
Aprod	81 (6)	100 (0)		71 (5)	50 (7)	73 (9)	67 (14)	83 (7)	70 (5)	89 (2)	63 (14)	90 (1)	88 (1)	48 (4)	90 (2)	72 (7)	63 (7)				
n	112	5		246	166	203	147	93	171	489	14	964	692	258	419	42	79				4100

approximately 82% when agreement also allowed a match with one of the alternate reference change labels. Producer's accuracies were typically lower than user's accuracies, indicating high change omission error. Producer's accuracies of change were 24.4%–30.3% for agreement defined as a match with the primary change reference label only, and increased to approximately 44.6%–47.2% when agreement also allowed a match with the alternate reference change labels. Overall accuracies for binary change classification tended to be higher by 0.8%–2.5% in the western sampling region than the eastern sampling region because of higher accuracies for the no change class (Table 18–23). User's accuracies for binary change tended to be higher in the eastern sampling region when the definition of agreement was defined as a match between the map label and primary reference label only, but were essentially equivalent when the alternate reference label was included in the definition of agreement. Producer's accuracies tended to be distinctly higher (>10%) in the eastern sampling region than the west regardless of agreement definition.

Consistent with the agreement statistics reported for the binary change and no change classification, agreement for the change reporting themes was generally poor (Table 24). Only the user's accuracies for forest loss was consistently near 80% for the three NLCD change periods. Urban gain user's accuracy approached 80% for the 2001–2011 and 2001–2006 change periods, but dropped to 68% for the 2006–2011 change period. Forest gain user's accuracies were between 71% and 74% for all three NLCD change periods. User's accuracies for most of the remaining reporting themes ranged from 50% to 70% with agriculture gain and water gain being exceptions with user's accuracies below 50%. Producer's accuracies for the change reporting themes were commonly below 50%. There was some regional differentiation in the user's accuracies for forest loss and forest gain (Tables 25, 26), with higher user's accuracies for forest loss in the western sampling region and higher user's accuracies for forest gain in the eastern sampling region.

In contrast to the change reporting themes, the no change reporting themes had higher agreement (Table 24). User's accuracies for all three NLCD time periods were >85% for four of the six no change reporting themes, and >80% for all no change reporting themes. Producer's accuracies for the six no change reporting themes exceeded 70% except urban. There was a stark regional difference in the user's and producer's accuracies for the shrubland no change and grassland no change reporting themes between east and west regions (Tables 25, 26). User's accuracies for shrubland no change and grassland no change exceeded 85% in the western region, but were 30% or less in the eastern region. Similarly, producer's accuracies for shrubland no change were about 60% higher in the west region than the east region, and producer's accuracies for grassland no change were approximately 10% higher in the west region than the east region. Conversely, user's accuracies for urban and forest no change tended to be approximately 5% higher in the east region than the west region.

4. Discussion

4.1. Comparison of NLCD 2011 accuracy assessment methods with “good practice” recommendations

The sampling design, response design, and analysis protocols implemented in the NLCD 2011 closely match the “good practice” recommendations for accuracy assessment described by Olofsson et al. (2014). Throughout the entirety of the NLCD program dating back to the accuracy assessment of NLCD 1992, probability sampling designs have been the basis for applying rigorous design-based inference (Stehman, 2000) to serve as the scientific foundation of the accuracy estimates and standard errors (Stehman et al., 2003; Wickham et al., 2004, 2010, 2013). The NLCD 2011 assessment continued to meet this “good practice” recommendation as we implemented a stratified random sampling design for collecting reference data. Our sampling design also followed

Table 9

Agreement between map and reference labels for NLCD 2011 for the continental United States at Level I of the classification hierarchy. See Table 4 for explanation of contents. $OA_1 = 74.5\%$ ($\pm 0.6\%$) and $OA_2 = 88.0$ ($\pm 0.4\%$).

Map ↓	Reference									User	Auser	n
	10	20	30	40	50	70	80	90	Total			
10	1.5720	0.0432	0.0137	0.0164	0.0079	0.0052	0.0171	0.1002	1.7757	89 (2)	91 (2)	214
20	0.0065	4.2506	0.0304	0.5230	0.1319	0.1756	0.6999	0.0412	5.8143	72 (2)	84 (2)	1758
30	0.0531	0.0274	0.5486	0.0233	0.2476	0.2876	0.0119	0.0287	1.2282	45 (4)	60 (4)	244
40	0.0362	0.5392	0.0098	21.8308	1.9046	0.3243	0.2810	0.4008	25.3262	86 (1)	94 (1)	1712
50	0.0501	0.3776	0.1182	2.0280	15.4971	3.9180	0.3795	0.0420	22.4105	69 (2)	88 (1)	1224
70	0.0388	0.3481	0.1034	0.6566	3.5890	7.9595	2.3091	0.0785	15.1190	53 (2)	81 (2)	1022
80	0.0685	0.9993	0.0173	0.9767	0.3127	0.6027	19.7591	0.2901	23.0263	86 (1)	92 (1)	1337
90	0.1185	0.0910	0.0044	1.3853	0.2564	0.1411	0.1205	3.1736	5.2998	60 (3)	75 (2)	489
Total	1.9438	6.6675	0.8457	27.4396	21.9563	13.4138	23.5781	4.1552	100.0000			
Prod	81 (3)	63 (2)	65 (7)	80 (1)	71 (1)	59 (2)	76 (2)	76 (2)				
Aprod	86 (3)	80 (2)	81 (6)	88 (1)	90 (1)	88 (1)	91 (2)	91 (2)				
n	232	1623	143	2108	1198	857	1461	378				8000

Table 10

Agreement between map and reference labels for NLCD 2006 for the continental United States at Level I of the classification hierarchy. See Table 4 for explanation of contents. $OA_1 = 75.2\%$ ($\pm 0.6\%$) and $OA_2 = 89.0$ ($\pm 0.4\%$).

Map ↓	Reference									User	Auser	n
	10	20	30	40	50	70	80	90	Total			
10	1.5314	0.0315	0.0216	0.0235	0.0079	0.0133	0.0205	0.0918	1.7415	88 (2)	92 (2)	206
20	0.0061	4.1046	0.0296	0.4411	0.2003	0.1839	0.7158	0.0407	5.7222	72 (2)	83 (2)	1156
30	0.0456	0.0187	0.5558	0.0309	0.2492	0.2863	0.0078	0.0287	1.2231	45 (4)	61 (4)	243
40	0.0438	0.4578	0.0146	22.1591	1.9763	0.3788	0.3113	0.4159	25.7576	86 (1)	95 (1)	2027
50	0.0521	0.3904	0.1231	1.6755	15.8257	3.7683	0.3562	0.0441	22.2354	71 (2)	89 (1)	1305
70		0.3961	0.1024	0.6019	3.5237	7.9991	2.2790	0.0693	14.9714	53 (2)	82 (2)	1231
80	0.0933	0.9728	0.0046	0.8914	0.3050	0.6393	19.8632	0.2778	23.0474	86 (1)	93 (1)	1343
90	0.1149	0.0912	0.0044	1.3606	0.2636	0.1628	0.1164	3.1874	5.3013	60 (3)	76 (2)	489
Total	1.8872	6.4631	0.8561	27.1840	22.3518	13.4319	23.6703	4.1557	100.0000			
Prod	81 (3)	64 (2)	65 (7)	82 (1)	71 (1)	60 (2)	84 (1)	77 (2)				
Aprod	87 (3)	81 (2)	82 (6)	90 (1)	90 (1)	87 (1)	90 (1)	92 (1)				
n	220	1388	140	2227	1257	859	1525	384				8000

Table 11

Agreement between map and reference labels for NLCD 2001 for the continental United States at Level I of the classification hierarchy. See Table 4 for explanation of contents. $OA_1 = 75.8$ ($\pm 0.6\%$) and $OA_2 = 89.3\%$ ($\pm 0.4\%$).

Map ↓	Reference									User	Auser	n
	10	20	30	40	50	70	80	90	Total			
10	1.5732	0.0315	0.0214	0.0281	0.0079	0.0131	0.0079	0.0997	1.7829	88 (2)	92 (2)	216
20	0.0054	3.9146	0.0276	0.4879	0.1628	0.1816	0.7500	0.0404	5.5703	70 (2)	82 (2)	835
30	0.0431	0.0287	0.5334	0.0179	0.2455	0.2791	0.0078	0.0287	1.1843	45 (4)	62 (2)	234
40	0.0438	0.4077	0.0098	22.7164	1.9762	0.2875	0.3083	0.4191	26.1687	87 (1)	95 (1)	2212
50	0.0480	0.3647	0.0784	1.6001	15.8047	3.7531	0.3428	0.0541	22.0459	72 (2)	90 (1)	1365
70	0.0002	0.3065	0.1075	0.5988	3.4782	8.0118	2.2633	0.0659	14.8322	54 (2)	82 (2)	1228
80	0.0644	0.9355	0.0046	0.9248	0.2622	0.6341	20.0202	0.2721	23.1180	87 (1)	93 (1)	1422
90	0.1100	0.0918	0.0044	1.4163	0.1727	0.1603	0.1259	3.2163	5.2978	61 (3)	77 (2)	488
Total	1.8882	6.0809	0.7874	27.7904	22.1104	13.3206	23.8263	4.1963	100.0000			
Prod	83 (3)	64 (2)	68 (7)	82 (1)	72 (1)	60 (2)	84 (1)	77 (2)				
Aprod	89 (3)	83 (2)	86 (6)	89 (1)	91 (1)	88 (1)	91 (1)	91 (2)				
n	222	1060	127	2507	1203	857	1638	386				8000

the “good practice” recommendations of stratifying by map class to reduce standard errors of accuracy estimates for the rare change types as well as rare land-cover classes, stratifying by subregions (east and west) to reduce standard errors of sub-region specific estimates, and implementing a simple random selection protocol within each stratum to allow unbiased estimation of variance of the accuracy estimates. Because cluster sampling would not have yielded substantial cost savings, we did not use clusters in the sampling design. Previous NLCD assessments did use clusters because at the time these assessments were implemented there were substantial savings in using clusters, as for example in the NLCD 1992 assessment (Stehman et al., 2003; Wickham et al., 2004) when hard-copy aerial photographs were used to determine the reference class.

Our analysis protocol follows the “good practice” recommendations (Olofsson et al., 2014, Sec. 6.4) nearly verbatim. Error matrices are reported in terms of proportion of area, we estimate user's and producer's accuracies for each class, the estimators are unbiased, we quantify variability by reporting standard errors, we use design-based inference, and we assess the impact of reference data uncertainty by reporting results for two definitions of agreement (i.e., with and without a match to the alternate reference labels). The primary difference from the “good practice” recommendations is that we do not emphasize in our reporting the area estimates based on the reference classification. The primary objectives of the NLCD 2011 assessment focus on documenting the accuracy of the single-date and change products to inform users of NLCD 2011 data in their applications. While the error matrices we report include

Table 12

Agreement between map and reference labels for NLCD 2011 for the eastern United States at Level I of the classification hierarchy. See Table 4 for explanation of contents. $OA_1 = 78.2 (\pm 0.8\%)$; $OA_2 = 87.0\% (\pm 0.6\%)$.

Map ↓	Reference										User	Auser	n
	10	20	30	40	50	70	80	90	Total				
10	2.3453	0.0873		0.0405			0.0228	0.2379			86 (3)	89 (3)	100
20	0.0149	6.7616	0.0448	1.0841	0.0410	0.0091	1.2241	0.1019	9.2816		73 (3)	85 (2)	891
30	0.0479	0.0365	0.1016	0.0365	0.0210	0.0365	0.0193	0.0077	0.3070		33 (6)	43 (7)	110
40	0.0896	1.2081		34.1965	0.7301	0.2717	0.6707	0.8929	38.0596		90 (1)	96 (1)	1045
50	0.0074	0.3523		2.1738	0.6731	0.3189	0.2970	0.1035	3.9260		17 (2)	28 (3)	399
70		0.2763	0.0024	0.8336	0.3525	0.5944	1.0036	0.0780	3.1408		19 (3)	39 (4)	346
80	0.1595	1.7085	0.0415	2.1411	0.3661	0.1953	26.6177	0.3770	31.6068		84 (1)	92 (1)	721
90	0.1647	0.1815		3.1138	0.4109	0.0794	0.1267	6.8673	10.9443		63 (3)	78 (3)	288
Total	2.8294	10.6122	0.1903	43.6198	2.5920	1.5052	29.9819	8.6663	100.0000				
Prod	83 (4)	64 (3)	53 (15)	78 (1)	26 (4)	40 (6)	89 (1)	79 (3)					
Aprod	88 (4)	80 (2)	63 (15)	91 (1)	48 (5)	67 (6)	93 (1)	93 (2)					
n	115	861	50	1434	234	165	784	257					3900

Table 13

Agreement between map and reference labels for NLCD 2011 for the western United States at Level I of the classification hierarchy. See Table 4 for explanation of contents. $OA_1 = 72.1 (\pm 1.0\%)$; $OA_2 = 89.2\% (\pm 0.6\%)$.

Map ↓	Reference										User	Auser	n
	10	20	30	40	50	70	80	90	Total				
10	1.0475	0.0133	0.0230		0.0133	0.0087	0.0133	0.0068	1.1260		93 (2)	95 (2)	114
20	0.0009	2.4724	0.0206	0.1426	0.1936	0.2886	0.3445		3.4631		71 (3)	81 (3)	867
30	0.0567	0.0212	0.8517	0.0143	0.4012	0.4579	0.0068	0.0430	1.8528		46 (4)	62 (4)	134
40		0.0856	0.0164	13.4445	2.7010	0.3599	0.0168	0.0672	16.6914		81 (2)	92 (1)	667
50	0.0791	0.3947	0.1984	1.9292	25.5495	6.3586	0.4354	0.0004	34.9542		73 (2)	92 (1)	825
70	0.0651	0.4573	0.1720	0.5366	5.7838	12.9538	3.1943	0.0788	23.2416		56 (2)	85 (2)	676
80	0.0068	0.5184	0.0008	0.1871	0.2765	0.8789	15.1081	0.2312	17.2078		88 (1)	93 (1)	616
90	0.0871	0.0296	0.0074	0.2132	0.1667	0.1829	0.1163	0.6688	1.4721		45 (3)	62 (3)	201
Total	1.3432	3.9925	1.2902	16.4675	35.0856	21.4893	19.2355	1.0961	100.0000				
Prod	78 (6)	62 (4)	66 (8)	82 (2)	73 (1)	60 (2)	79 (2)	61 (6)					
Aprod	82 (6)	79 (4)	83 (7)	91 (1)	92 (1)	89 (1)	87 (2)	81 (5)					
n	117	762	93	674	964	692	677						4100

Table 14

Overall accuracies for each NLCD era by agreement definition, sampling region, and map homogeneity. The label Pri represents agreement based on a match between the map and primary reference labels and PriAlt represents agreement based on a match between the map and either the primary or alternate reference labels. Homogeneous subset had 1811 and 2185 sample pixels in east and west, respectively.

Year	CONUS		EAST		WEST	
	Pri	PriAlt	Pri	PriAlt	Pri	PriAlt
Level 2						
2011	66	82	63	76	68	86
2006	67	83	64	77	69	87
2001	67	83	64	78	69	87
Level 1						
2011	75	88	78	87	72	89
2006	75	89	79	88	73	90
2001	76	89	80	89	73	90
Homogeneous subset						
Level 2						
2011	75	89	74	85	76	91
2006	75	89	74	85	76	92
2001	75	89	74	86	76	91
Level 1						
2011	84	94	91	95	80	94
2006	84	95	91	95	80	94
2001	85	95	91	96	80	94

the estimated percent of area of each class (based on the reference classification), it is not a primary intent of the NLCD program to produce these area estimates.

The response design protocol also follows the “good practice” guidelines very closely. The reference data provided the required temporal representation consistent with the change period of the map, we assigned each pixel a primary and secondary (if warranted) reference label to account for uncertainty in the labeling protocol, and the response design included several procedures to ensure interpreter

consistency. The one “good practice” suggestion we did not include was that we did not collect interpreter confidence ratings for each pixel. We had collected confidence ratings in previous NLCD assessments but found that interpreters had difficulty being consistent when assigning these confidence ratings. Analyses showed that interpreter confidence was not as strongly associated with classification error as features such as the complexity of the landscape surrounding the sample pixel (Wickham et al., 2010) so we decided not to burden the interpreters with this extra requirement of a confidence rating.

Table 15

Agreement between map and reference labels for binary change versus no change for 2001–2011 for the continental United States. See Table 4 for explanation of contents. $OA_1 = 95.2\%$ ($\pm 0.3\%$) and $OA_2 = 97.8\%$ ($\pm 0.2\%$).

		Reference					n
		NoChange	Change	Total	User	Auser	
Map	NoChange	93.704	3.505	97.209	96.4 (0.3)	98.3 (0.2)	5339
	Change	1.269	1.521	2.791	54.5 (1.4)	82.0 (1.7)	2661
	Total	94.973	5.026				
	Prod	98.7 (0.04)	30.3 (1.8)				
	Aprod	99.0 (0.04)	47.2 (2.7)				
	n	6326	1674				

Table 16

Agreement between map and reference labels for binary change versus no change for 2006–2011 for the continental United States. See Table 4 for explanation of contents. $OA_1 = 96.6\%$ ($\pm 0.2\%$) and $OA_2 = 98.8\%$ ($\pm 0.2\%$).

		Reference					n
		NoChange	Change	Total	User	Auser	
Map	NoChange	95.692	2.629	98.321	99.0 (0.2)	97.3 (0.2)	6213
	Change	0.728	0.951	1.679	56.7 (1.8)	82.6 (1.3)	1787
	Total	96.420	3.580				
	Prod	99.2 (0.03)	26.6 (1.9)				
	Aprod	99.4 (0.03)	47.2 (3.4)				
	n	6885	1115				

Table 17

Agreement between map and reference labels for binary change versus no change for 2001–2006 for the continental United States. See Table 4 for explanation of contents. $OA_1 = 96.6\%$ ($\pm 0.2\%$) and $OA_2 = 98.2\%$ ($\pm 0.15\%$).

		Reference					n
		NoChange	Change	Total	User	Auser	
Map	NoChange	95.728	2.673	98.401	97.3 (0.2)	99.1 (0.2)	6961
	Change	0.734	0.864	1.598	54.1 (2.0)	82.9 (1.7)	1039
	Total	96.462	3.537				
	Prod	99.2 (0.03)	24.4 (1.8)				
	Aprod	99.4 (0.03)	44.6 (3.1)				
	n	6945	1055				

Table 18

Agreement between map and reference labels for binary change versus no change for 2001–2011 for the eastern United States. See Table 4 for explanation of contents. $OA_1 = 93.7\%$ (0.4%) and $OA_2 = 97.1\%$ (0.3%).

		Reference					n
		NoChange	Change	Total	Users	Auser	
Map	NoChange	91.255	4.472	95.727	95.3 (0.4)	97.8 (0.3)	
	Change	1.784	2.488	4.272	58.2 (1.9)	81.7 (1.5)	
	Total	93.039	6.960				
	Prod	98.1 (0.1)	35.7 (2.2)				
	AProd	99.2 (0.1)	62.5 (3.2)				
	n						

Table 19

Agreement between map and reference labels for binary change versus no change for 2001–2011 for the western United States. See Table 4 for explanation of contents. $OA_1 = 96.2\%$ (0.4%) and $OA_2 = 98.3\%$ (0.3%).

		Reference					n
		NoChange	Change	Total	Users	Auser	
Map	NoChange	95.365	2.850	98.215	97.1 (0.4)	98.6 (0.3)	
	Change	0.920	0.866	1.786	48.5 (2.0)	82.3 (1.8)	
	Total	96.285	3.716				
	Prod	99.0 (0.1)	23.3 (2.5)				
	AProd	99.7 (0.03)	51.3 (4.8)				
	n						

4.2. Accuracy of NLCD 2011 land cover

The approximate 83% overall accuracies of the single-date maps for all NLCD eras at the 16-class (Level II) hierarchical level approached

the nominal 85% quality benchmark, and 6 of the 16 classes (water, high density urban, deciduous forest, evergreen forest, and shrubland) had user's accuracies that met or exceeded the nominal benchmark. At the 8-class (Level I) hierarchical level, overall accuracies for all NLCD

Table 20

Agreement between map and reference labels for binary change versus no change for 2006–2011 for the eastern United States. See Table 4 for explanation of contents. $OA_1 = 95.4\%$ (0.4%) and $OA_2 = 98.3\%$ (0.2%).

		Reference		Total	Users	Auser
		NoChange	Change			
Map	NoChange	93.710	3.448	97.158	96.5 (0.4)	98.8 (0.2)
	Change	1.143	1.700	2.843	59.8 (2.3)	82.5 (1.7)
	Total	94.853	5.148			
	Prod	98.8 (0.1)	33.0 (2.4)			
	AProd	99.5 (0.04)	65.9 (4.2)			

Table 21

Agreement between map and reference labels for binary change versus no change for 2006–2011 for the western United States. See Table 4 for explanation of contents. $OA_1 = 97.5\%$ (0.3%) and $OA_2 = 99.1\%$ (0.2%).

		Reference		Total	Users	Auser
		NoChange	Change			
Map	NoChange	97.037	2.073	99.110	97.9 (0.3)	99.2 (0.2)
	Change	0.446	0.444	0.890	49.9 (2.8)	82.7 (2.0)
	Total	97.483	2.517			
	Prod	99.5 (0.1)	17.7 (2.4)			
	AProd	99.8 (0.02)	49.1 (6.4)			

Table 22

Agreement between map and reference labels for binary change versus no change for 2001–2006 for the eastern United States. See Table 4 for explanation of contents. $OA_1 = 95.3\%$ (0.4%) and $OA_2 = 98.3\%$ (0.2%).

		Reference		Total	Users	Auser
		NoChange	Change			
Map	NoChange	93.852	3.847	97.699	96.1 (0.4)	98.6 (0.2)
	Change	0.866	1.435	2.301	62.4 (2.6)	83.6 (2.2)
	Total	94.718	5.282			
	Prod	99.1 (0.1)	27.2 (2.1)			
	AProd	99.6 (0.1)	59.2 (4.1)			

Table 23

Agreement between map and reference labels for binary change versus no change for 2001–2006 for the western United States. See Table 4 for explanation of contents. $OA_1 = 97.5\%$ (0.3%) and $OA_2 = 99.2\%$ (0.2%).

		Reference		Total	Users	Auser
		NoChange	Change			
Map	NoChange	97.000	1.877	98.877	98.1 (0.3)	99.4 (0.2)
	Change	0.645	0.478	1.123	42.5 (3.0)	81.9 (2.6)
	Total	97.645	2.355			
	Prod	99.3 (0.1)	20.3 (2.8)			
	AProd	99.8 (0.02)	59.9 (6.5)			

eras were 88% or higher, exceeding the nominal 85% quality benchmark, and high user's accuracies ($\geq 85\%$) were realized for water, urban, forest, shrubland, and agriculture.

Ranging from 33% to 93% across the three change eras, the emergent pattern across the three change eras was high user's accuracies for no change reporting themes, urban gain, and forest loss and gain. The remaining change reporting themes had lower user's accuracies. A partial explanation for the lack of uniformly high user's accuracies for reporting themes representing change is evident in the error matrices. Approximately 14% of the Level I disagreement is attributable to map-reference mismatches between forest (class 40) and shrubland (class 50), shrubland and grassland (class 70), and grassland and agriculture (class 80). Disagreement among these classes suggests that determination of the most appropriate class label at "interfaces" across the forest-shrubland-grassland gradient is difficult, and, likewise, determination of the context of grassland-dominated areas (grassland, agriculture, open urban (class 21)) is difficult at the mapping phase, reference label assignment phase, or both. Less disagreement among these classes likely would have led to improved agreement across the loss and gain reporting themes. A portion of the disagreement among these classes

is also likely attributable to the inherent ambiguity in class definitions (Lunetta et al., 2001; Mann and Rothley, 2006).

Several researchers and previous NLCD accuracy assessments have shown that map accuracy tends to improve in areas that are homogeneously classified (Löw et al., 2015; Smith et al., 2002, 2003; van Oort et al., 2004; Wickham et al., 2010, 2013; Yu et al., 2008). In other words, map-reference agreement tends to be more likely when neighboring pixels have the same map label as the sample pixel. The positive relationship between map homogeneity and agreement reported in previous assessments was also found in this assessment. The relationship between map homogeneity and agreement suggests that user's and producer's accuracies for the 11 loss and gain reporting themes are probably higher for larger, more homogeneous areas of change and lower for smaller areas of change (e.g., single, isolated pixels) than reported in Table 14.

4.3. Comparison of NLCD 2011 and NLCD 2006 accuracies

The agreement statistics reported here for year 2006, year 2001, and the 2001–2006 change reporting themes can be compared to their

Table 24

User's and Producer's accuracies (%) for reporting themes for the continental United States. Standard errors are in parentheses. The symbol Δ = change. Agreement is defined as a match between the map labels and either the primary or alternate reference labels.

Theme	User's accuracy			Producer's accuracy		
	2001–2011	2006–2011	2001–2006	2001–2011	2006–2011	2001–2006
Water loss	65 (11)	45 (17)	86 (8)	60 (13)	29 (17)	63 (11)
Water gain	61 (12)	87 (8)	36 (15)	32 (11)	42 (12)	19 (10)
Urban gain	79 (2)	68 (3)	78 (3)	30 (4)	23 (5)	28 (5)
Forest loss	82 (2)	79 (2)	80 (3)	51 (3)	54 (5)	37 (3)
Forest gain	74 (3)	72 (4)	71 (5)	22 (2)	19 (3)	21 (3)
Shrub loss	58 (3)	59 (4)	60 (5)	20 (2)	16 (2)	17 (2)
Shrub gain	62 (2)	64 (3)	63 (4)	35 (3)	30 (3)	23 (3)
Grass loss	54 (4)	61 (4)	57 (5)	20 (3)	21 (3)	18 (3)
Grass gain	59 (3)	67 (3)	72 (4)	33 (4)	33 (5)	29 (4)
Ag loss	55 (5)	66 (7)	49 (6)	26 (5)	26 (7)	27 (6)
Ag gain	38 (7)	47 (9)	33 (9)	24 (7)	25 (10)	25 (9)
Water no Δ	89 (2)	90 (2)	90 (2)	82 (3)	82 (3)	83 (3)
Urban no Δ	82 (2)	83 (2)	82 (2)	68 (2)	67 (2)	68 (2)
Forest no Δ	93 (1)	93 (1)	94 (1)	82 (1)	82 (1)	83 (1)
Shrub no Δ	88 (1)	88 (1)	89 (1)	77 (1)	77 (1)	77 (1)
Grass no Δ	82 (2)	81 (2)	82 (2)	72 (2)	71 (2)	72 (2)
Ag no Δ	92 (1)	92 (1)	93 (1)	85 (1)	85 (1)	85 (1)

Table 25

User's and Producer's accuracies (%) and standard errors (in parentheses) for the eastern United States. All values are rounded to the nearest integer. The symbol Δ = change. Agreement is defined as a match between the map labels and either the primary or alternate reference labels.

Theme	User's accuracy			Producer's accuracy		
	2001–2011	2006–2011	2001–2006	2001–2011	2006–2011	2001–2006
Water loss	63 (17)	67 (27)	80 (18)	58 (22)	41 (27)	80 (18)
Water gain	60 (16)	67 (19)	25 (22)	49 (18)	99 (1)	12 (11)
Urban gain	78 (3)	68 (4)	77 (4)	32 (5)	20 (6)	39 (7)
Forest loss	80 (2)	76 (3)	81 (3)	54 (4)	59 (5)	36 (4)
Forest gain	75 (4)	73 (4)	73 (6)	26 (3)	23 (3)	23 (3)
Shrub loss	55 (4)	59 (4)	62 (6)	21 (3)	17 (3)	16 (2)
Shrub gain	56 (3)	60 (4)	69 (5)	35 (4)	36 (4)	20 (3)
Grass loss	54 (5)	62 (5)	56 (6)	24 (4)	29 (5)	19 (4)
Grass gain	52 (4)	61 (5)	71 (5)	46 (6)	45 (6)	38 (5)
Ag loss	58 (6)	65 (9)	49 (8)	27 (6)	23 (9)	26 (7)
Ag gain	33 (13)	39 (18)	29 (17)	18 (9)	21 (13)	15 (11)
Water no Δ	88 (3)	89 (3)	91 (3)	83 (4)	82 (4)	84 (4)
Urban no Δ	84 (2)	85 (2)	84 (2)	69 (2)	69 (2)	69 (2)
Forest no Δ	95 (1)	95 (1)	95 (1)	81 (1)	81 (1)	82 (1)
Shrub no Δ	10 (3)	13 (3)	26 (4)	29 (9)	27 (6)	45 (6)
Grass no Δ	32 (5)	32 (4)	32 (5)	68 (8)	66 (7)	66 (8)
Ag no Δ	92 (1)	92 (1)	92 (1)	90 (1)	90 (1)	90 (1)

Table 26

User's and Producer's accuracies (%) and standard errors (in parentheses) for the western United States. All values are rounded to the nearest integer. The symbol Δ = change. Agreement is defined as a match between the map labels and either the primary or alternate reference labels.

Theme	User's accuracy			Producer's accuracy		
	2001–2011	2006–2011	2001–2006	2001–2011	2006–2011	2001–2006
Water loss	67 (14)	33 (19)	88 (8)	61 (17)	22 (19)	59 (12)
Water gain	63 (17)	100 (0)	43 (19)	22 (12)	34 (12)	25 (17)
Urban gain	81 (2)	67 (5)	78 (4)	27 (7)	31 (13)	18 (5)
Forest loss	87 (2)	86 (3)	78 (5)	47 (7)	46 (9)	39 (8)
Forest gain	71 (5)	59 (18)	54 (7)	7 (2)	4 (2)	10 (4)
Shrub loss	62 (4)	57 (6)	58 (6)	19 (4)	13 (3)	18 (4)
Shrub gain	71 (4)	73 (5)	57 (6)	35 (6)	23 (5)	27 (6)
Grass loss	54 (5)	59 (5)	57 (7)	17 (4)	13 (3)	16 (5)
Grass gain	68 (4)	76 (3)	74 (5)	26 (5)	25 (5)	21 (4)
Ag loss	48 (9)	66 (11)	49 (11)	24 (7)	30 (12)	32 (12)
Ag gain	40 (8)	51 (9)	35 (10)	27 (10)	27 (15)	31 (13)
Water no Δ	92 (3)	92 (3)	90 (3)	81 (5)	81 (5)	81 (5)
Urban no Δ	79 (3)	80 (3)	79 (3)	66 (4)	64 (4)	66 (4)
Forest no Δ	90 (2)	91 (1)	92 (1)	85 (2)	85 (2)	85 (2)
Shrub no Δ	91 (1)	92 (1)	93 (1)	77 (1)	78 (1)	78 (1)
Grass no Δ	85 (2)	85 (2)	85 (2)	72 (2)	71 (2)	72 (2)
Ag no Δ	93 (1)	93 (1)	93 (1)	80 (2)	80 (2)	80 (2)

counterparts from the NLCD 2006 accuracy assessment (Wickham et al., 2013). The Level II and Level I overall accuracies for the single-date assessments for 2006 and 2001 reported here (about 82% and 88%,

respectively) were approximately 4% greater than their counterparts for the assessment of the NLCD 2006 product. The improvements in NLCD 2011 overall accuracies were modest but significant since the

standard errors for all overall accuracies reported here and in the NLCD 2006 assessment were <1%. The improved overall accuracies for both hierarchical levels of NLCD 2011 are primarily attributable to improved user's accuracies for low density urban (class 22), medium density urban (class 23), woody wetland (90), and emergent wetland (95). User's accuracies for the two urban classes and two wetland classes were approximately 10% and 30% higher, respectively, for the NLCD 2011 product than for NLCD 2006 product. User's accuracies for perennial snow and ice (class 12), mixed forest (43) and pasture (class 81) were higher in the NLCD 2006 product than the NLCD 2011 product, but the lower user's accuracies for these classes in the NLCD 2011 product did not affect NLCD 2011 Level II or Level I overall accuracies, which were higher than their NLCD 2006 counterparts. Among both the static and dynamic 2001–2006 change reporting themes, the NLCD 2011 product had higher user's accuracies for urban gain (NLCD 2011: $78\% \pm 3\%$; NLCD 2006: $72\% \pm 1\%$), urban–no change (NLCD 2011: $82\% \pm 2\%$; NLCD 2006: $73\% \pm 2\%$), shrubland–no change (NLCD 2011: $89\% \pm 1\%$; NLCD 2006: $85\% \pm 2\%$), and grassland–no change (NLCD 2011: $82\% \pm 2\%$; NLCD 2006: $75\% \pm 3\%$). User's accuracies for most of the other change reporting themes were statistically equivalent, and statistical equivalence may have been partly attributable to higher standard errors for NLCD 2011 in some cases. For example, user's accuracy for the 2001–2006 water loss theme was $86\% \pm 8\%$ for the NLCD 2011 product and $80\% \pm 2\%$ for the NLCD 2006 product. The approximate 50% reduction in the number of sample pixels for NLCD 2011 accuracy assessment compared to the NLCD 2006 accuracy assessment contributed to the higher standard errors. The change reporting themes of shrubland gain and agriculture loss and gain were other examples of statistical equivalence that may have been attributable to high standard errors for the NLCD 2011 accuracy assessment. The user's accuracy for change in the binary change–no change reported here ($82.9\% \pm 1.7\%$; Table 17) was about equivalent to its counterpart in the NLCD 2006 assessment ($84.5\% \pm 0.6\%$).

4.4. Comparison of NLCD 2011 to other land cover change efforts

More recently there has been an emphasis on accuracy assessment of land cover change because of the wide ranging impacts of land cover change on biodiversity, carbon dynamics, water quality, and other aspects of environmental condition. The user's and producer's accuracies reported here for forest loss, forest gain, and urban gain compare favorably with recent land cover change accuracy assessments. On average, our continental forest gain and forest loss user's accuracies were 30% to 35% higher than forest gain and forest loss user's accuracies for the temperate forest biome reported by Feng et al. (2016, p. 80), and approximately 23% higher than those reported for temperate forests by Potapov et al. (2011, p. 557). The producer's accuracies reported for forest loss and forest gain by Feng et al. (2016) were 6%–9% higher than NLCD 2011, and forest loss producer's accuracy reported by Potapov et al. (2011) was approximately 13% higher than NLCD 2011. Yuan et al. (2005) reported a user's accuracy of 66% across all types of change in metropolitan Minneapolis, Minnesota (USA), which is about 10% lower than our urban gain user's accuracies for 2001–2006 and 2001–2011 change periods. The NLCD 2011 products of year 2001 (version 3), 2006 (version 2) and NLCD 2011 (version 1), when used in tandem, appear to provide accurate data for determining where urbanization has occurred, where forests have changed, and where land cover has not changed.

Acknowledgements

U.S. Environmental Protection Agency, through its Office of Research and Development, partly funded and managed the research described here. The article has been reviewed by the USEPA's Office of Research

and Development and approved for publication. Approval does not signify that the contents reflect the views of the USEPA. S. Stehman's participation was underwritten by contract G12AC20221 between SUNY-ESF and USGS.

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. Geological Survey, Geological Survey Professional Paper 964. http://www.pbcgis.com/data_basics/anderson.pdf.
- Cochran, W.G., 1977. Sampling Techniques. third ed. John Wiley & Sons, New York.
- Comber, A., Fisher, P., Wadsworth, R., 2005. What is land cover? Environ. Plann. B Plann. Des. 32, 199–209.
- Coulston, J.W., Moisen, G.G., Wilson, B.T., Finco, M.V., Cohen, W.B., Brewer, C.K., 2012. Modeling percent tree canopy cover. Photogramm. Eng. Remote. Sens. 78, 715–727.
- Feng, M., Sexton, J.O., Huang, C., Anand, A., Channan, S., Song, X.-P., Song, D.-X., Kim, D.-H., Noojipady, P., Townshend, J.R., 2016. Earth science data records of global forest cover and change: assessment of accuracy in 1990, 2000, and 2005 epochs. Remote Sens. Environ. 184, 73–85.
- Foody, G.M., 2006. The evaluation and comparison of thematic maps derived from remote sensing. In: Caetano, M., Painho, M. (Eds.), Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 5–7 July 2006. Instituto Geográfico Português, Lisboa:pp. 18–31 (http://www.spatial-accuracy.org/system/files/Foody2006accuracy_0.pdf). Last accessed 12 December 2016).
- Fry, J.A., Xian, G., Jin, S., Dewitz, J.A., Homer, C.G., Yang, L., Barnes, C.A., Herold, N.D., Wickham, J., 2011. Completion of the 2006 National Land Cover Database for the conterminous United States. Photogramm. Eng. Remote. Sens. 77, 858–864.
- Gopal, S., Woodcock, C., 1994. Theory and methods for accuracy assessment of thematic maps using fuzzy sets. Photogramm. Eng. Remote. Sens. 60, 181–188.
- Homer, C., Gallant, A., 2001. Partitioning the conterminous United States in mapping zones for Landsat TM land cover mapping. USGS White Paper (<http://landcover.usgs.gov/pdf/homer.pdf>). Last accessed 12 December 2016).
- 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., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, N., Wickham, J., 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogramm. Eng. Remote. Sens. 73, 337–341.
- 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 National Land Cover Database for the conterminous United States—representing a decade of land cover change information. Photogramm. Eng. Remote. Sens. 81, 345–354.
- Löw, F., Knöfel, P., Conrad, C., 2015. Analysis of uncertainty in multi-temporal object-based classification. ISPRS J. Photogramm. Remote Sens. 105, 91–106.
- Lunetta, R.S., Liames, J., Knight, J., Congalton, R.G., Mace, T.H., 2001. An assessment of reference data variability using a “virtual field reference database”. Photogramm. Eng. Remote. Sens. 63, 707–715.
- Mann, S., Rothley, K.D., 2006. Sensitivity of Landsat/IKONOS accuracy comparison to errors in photointerpreted reference data and variations in test point sets. Int. J. Remote Sens. 27, 5027–5036.
- 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.
- Potapov, P., Turubanova, S., Hansen, M.C., 2011. Region-scale boreal forest cover and change mapping using Landsat data composites for European Russia. Remote Sens. Environ. 115, 548–561.
- Särndal, C.E., Swensson, B., Wretman, J., 1992. Model-Assisted Survey Sampling. Springer-Verlag, New York.
- Smith, J.H., Wickham, J.D., Stehman, S.V., Yang, L., 2002. Impacts of patch size and land cover heterogeneity on thematic image classification accuracy. Photogramm. Eng. Remote. Sens. 68, 65–70.
- Smith, J.H., Stehman, S.V., Wickham, J.D., Yang, L., 2003. Effects of landscape characteristics on land-cover class accuracy. Remote Sens. Environ. 84, 342–349.
- 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., Czaplewski, R.L., 1998. Design and analysis for thematic map accuracy assessment. Remote Sens. Environ. 64, 331–344.
- 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.
- Stehman, S.V., Wickham, J.D., Smith, J.H., Yang, L., 2003. Thematic accuracy of the 1992 National Land-Cover Data (NLCD) for the eastern United States: statistical methodology and regional results. Remote Sens. Environ. 86, 500–516.
- Stehman, S.V., Wickham, J., 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 United States. Photogramm. Eng. Remote. Sens. 74, 1561–1571.
- van Oort, P.A.J., Bregt, A.K., de Bruin, S., de Wit, A.J.W., Stein, A., 2004. Spatial variability in classification accuracy of agricultural crops in the Dutch national land-cover database. Int. J. Geogr. Inf. Sci. 18, 611–626.

- Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K., Van Driel, J.N., 2001. Completion of a 1990's National Land Cover Database for the conterminous United States. *Photogramm. Eng. Remote. Sens.* 67, 650–652.
- Wickham, J.D., Stehman, S.V., Smith, J.H., Yang, L., 2004. Thematic accuracy of the 1992 National Land-Cover Data for the western United States. *Remote Sens. Environ.* 91, 452–468.
- Wickham, J.D., Stehman, S.V., Fry, J.A., Smith, J.H., Homer, C.G., 2010. Thematic accuracy of the NLCD 2001 land cover for the conterminous United States. *Remote Sens. Environ.* 114, 1286–1296.
- Wickham, J., 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., Homer, C., Vogelmann, J., McKerrow, A., Mueller, R., Herold, N., Coulston, J., 2014. The Multi-Resolution Land Characteristics (MRLC) consortium—20 years of development and integration of USA National Land Cover Data. *Remote Sens.* 6: 7424–7441. <http://dx.doi.org/10.3390/rs6087424>.
- Xian, G., Homer, C., Dewitz, J., Fry, J., Hossain, N., Wickham, J., 2011. The change of impervious surface area between 2001 and 2006 in the conterminous United States. *Photogramm. Eng. Remote. Sens.* 67, 650–652.
- Yu, Q., Gong, P., Tian, Y.Q., Pu, R., Yang, J., 2008. Factors affecting spatial variation in classification uncertainty in an image object-based vegetation mapping. *Photogramm. Eng. Remote. Sens.* 77, 758–762.
- Yuan, F., Saway, K.E., Loeffelholz, B.C., Bauer, M.E., 2005. Land cover classification and change analysis of the Twin Cities (Minnesota) metropolitan area by multi-temporal Landsat remote sensing. *Remote Sens. Environ.* 98, 317–328.