Remote Sensing of Environment 115 (2011) 1401–1407

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Thematic accuracy of the National Land Cover Database (NLCD) 2001 land cover for Alaska

David J. Selkowitz [a,*], Stephen V. Stehman [b]

[a] U.S. Geological Survey, Alaska Science Center, 4210 University Drive, Anchorage AK, 99508, USA
[b] SUNY College of Environmental Science and Forestry, 320 Bray Hall, 1 Forestry Drive, Syracuse, New York, 13210, USA

ARTICLE INFO

Article history:
Received 24 June 2010
Received in revised form 17 January 2011
Accepted 29 January 2011

Keywords:
Accuracy assessment
Land cover classification
Alaska

ABSTRACT

The National Land Cover Database (NLCD) 2001 Alaska land cover classification is the first 30-m resolution land cover product available covering the entire state of Alaska. The accuracy assessment of the NLCD 2001 Alaska land cover classification employed a geographically stratified three-stage sampling design to select the reference sample of pixels. Reference land cover class labels were determined via fixed wing aircraft, as the high resolution imagery used for determining the reference land cover classification in the conterminous U.S. was not available for most of Alaska. Overall thematic accuracy for the Alaska NLCD was 76.2% (s.e. 2.8%) at Level II (12 classes evaluated) and 83.9% (s.e. 2.1%) at Level I (6 classes evaluated) when agreement was defined as a match between the map class and either the primary or alternate reference class label. When agreement was defined as a match between the map class and primary reference label only, overall accuracy was 59.4% at Level II and 69.3% at Level I. The majority of classification errors occurred at Level I of the classification hierarchy (i.e., misclassifications were generally to a different Level I class, not to a Level II class within the same Level I class). Classification accuracy was higher for more abundant land cover classes and for pixels located in the interior of homogeneous land cover patches.

Published by Elsevier Inc.

1. Introduction

The National Land Cover Database (NLCD) 2001 Alaska land cover classification is the first 30-m resolution land cover product available covering the entire state of Alaska (http://www.mrlc.gov). A circa 1992 NLCD land cover dataset has been available for the conterminous U.S. since 2000 (Vogelmann et al., 2001), with the second generation circa 2001 NLCD dataset available since 2007 (Homer et al., 2004, 2007). These products have been used extensively by researchers from a wide range of disciplines to study a variety of phenomena (see list of applications of NLCD provided by Stehman et al. (2008) and Wickham et al. (2010)). The NLCD Alaska land cover classification, available since 2008, fills a significant land cover data gap by providing coverage for many areas where no 30-m resolution (or finer) land cover data were previously available. Prior to the release of the NLCD 2001 Alaska dataset, the only existing statewide land cover classification product was a 1-km resolution AVHRR-derived dataset (Fleming, 1997) which was never subjected to a formal accuracy assessment. Alaska's vast size, limited availability of high resolution imagery, and unique land cover composition necessitated the development of a separate accuracy assessment methodology from that employed for the conterminous United States NLCD, detailed in Wickham et al. (2010). In this article, we document the Alaska accuracy assessment methodology and present the results obtained from the accuracy assessment.

2. Methods

The NLCD classification system is hierarchical, with broad Level I classes divided into more detailed Level II classes. The NLCD 2001 land cover product depicts 8 Level I and 19 Level II land cover classes across the state of Alaska, including three classes not mapped in the conterminous U.S. (Tables 1 and 2). The NLCD dataset was developed to provide a consistent land cover classification across all regions of the United States and meet the needs of a wide variety of users and applications. Consequently, NLCD classes were designed to provide a meaningful description of local and regional land cover variability while remaining consistent across the entire U.S. The NLCD 2001 product is derived from Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper-plus (ETM+) imagery and has a nominal spatial resolution of 30 m, a minimum mapping unit of 5 pixels, and is mapped to the Alaska Albers Equal Area Projection. NLCD 2001 is intended to represent land cover as close as possible to the year 2001; although the Landsat TM and ETM+ scenes used for mapping span the years 1999–2004, the vast majority of scenes were acquired between 1999 and 2002. Landsat scenes used to form the primary image mosaic were acquired between June and early September.

For the NLCD 2001 accuracy assessment for the conterminous U.S., image analysts used Digital Orthophoto Quarter Quadrangles (DOQQs)

* Corresponding author. Tel.: +1 907 786 7146.
E-mail address: dselkowitz@usgs.gov (D.J. Selkowitz).

0034-4257/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.rse.2011.01.020

Table 1
NLCD land cover class descriptions. NLCD classes not mapped in Alaska have been omitted.

11. Open water—All areas of open water, generally with less than 25 percent cover of vegetation or soil.
12. Perennial ice/snow—All areas characterized by a perennial cover of ice and/or snow, generally greater than 25% of total cover.
21. Developed, open space—Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20% of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes
22. Developed, low intensity—Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20–49% of total cover. These areas most commonly include single-family housing units.
23. Developed, medium intensity—Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50–79% of the total cover. These areas most commonly include single-family housing units.
24. Developed, high intensity—Includes highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses, and commercial/industrial. Impervious surfaces account for 80 to 100% of the total cover.
31. Barren land (rock/sand/clay)—Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits, and other accumulations of earthen material. Generally, vegetation accounts for less than 15% of total cover.
41. Deciduous forest—Areas dominated by trees generally greater than 5-m tall, and greater than 20% of total vegetation cover. More than 75% of the tree species shed foliage simultaneously in response to seasonal change.
42. Evergreen forest—Areas dominated by trees generally greater than 5-m tall, and greater than 20% of total vegetation cover. More than 75% of the tree species maintain their leaves all year. Canopy is never without green foliage.
43. Mixed forest—Areas dominated by trees generally greater than 5-M tall, and greater than 20% of total vegetation cover. Neither deciduous nor evergreen species are greater than 75% of total tree cover.
51. Dwarf scrub—Alaska only areas dominated by shrubs less than 20-cm tall with shrub canopy typically greater than 20% of total vegetation. This type is often co-associated with grasses, sedges, herbs, and non-vascular vegetation.
52. Shrub/scrub—Areas dominated by shrubs; less than 5-m tall with shrub canopy typically greater than 20% of total vegetation. This class includes true shrubs, young trees in an early successional stage, or trees stunted from environmental conditions.
71. Grassland/herbaceous—Areas dominated by grammanoid or herbaceous vegetation, generally greater than 80% of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing.
72. Sedge/herbaceous—Alaska only areas dominated by sedges and forbs, generally greater than 80% of total vegetation. This type can occur with significant other grasses or other grass like plants, and includes sedge tundra, and sedge tussock tundra.
74. Moss—Alaska only areas dominated by mosses, generally greater than 80% of total vegetation.
81. Pasture/hay—Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20% of total vegetation.
82. Cultivated crops—Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20% of total vegetation. This class also includes all land being actively tilled.
90. Woody wetlands—Areas where forest or shrubland vegetation accounts for greater than 20% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.
95. Emergent herbaceous wetlands—Areas where perennial herbaceous vegetation accounts for greater than 80% of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

or other high resolution imagery to provide reference land cover class labels. DOQQs or similar high resolution image products from circa 2001, however, are not available for most of Alaska. In addition, interpretation of high resolution imagery can be extremely difficult in areas where low-statured tundra vegetation is dominant. For these two reasons, it was necessary to adopt an alternative approach to accuracy assessment of the

Table 2
NLCD Anderson Level II classes, Level I class equivalents, class names, and percent area mapped in Alaska.

Class (Level II)	Level I class	Name	Percent of area
11	10	Open water	14.2
12	10	Perennial ice/snow	4.3
21	20	Developed, open space	0.02
22	20	Developed, low intensity	0.06
23	20	Developed, medium intensity	0.007
24	20	Developed, high intensity	0.003
31	30	Barren land	7.7
41	40	Deciduous forest	3.4
42	40	Evergreen forest	15.5
43	40	Mixed forest	3.5
51[a]	50	Dwarf scrub	17.2
52	50	Shrub/scrub	21.2
71	70	Grassland/herbaceous	0.8
72[a]	70	Sedge/herbaceous	5.7
74[a]	70	Moss	0.03
81	80	Pasture/hay	0.003
82	80	Cultivated crops	0.02
90	90	Woody wetlands	3.5
95	90	Emergent herbaceous wetlands	2.9

[a] Class not mapped in the conterminous U.S.

Alaska NLCD land cover classification. This section describes the major components of the methodology, the sampling design, response design, and analysis (Stehman & Czaplewski, 1998), used to assess the accuracy of the NLCD 2001 Alaska land cover classification.

2.1. Sampling design

The sampling design implemented for the NLCD 2001 Alaska accuracy assessment was a stratified three-stage cluster sample. The stratification was created by dividing the state of Alaska into 31 geographic regions (Fig. 1) corresponding roughly to the ecoregions defined by Nowacki et al. (2001). This geographic stratification was constructed to ensure that the sample was spatially well-distributed across Alaska. In addition, given the substantial resources required and logistical challenges inherent in such an undertaking, geographic stratification provided a structure that allowed the sample selection protocol to be modified (while still retaining the probability sampling character of the design) even if resource limitations or other challenges prohibited sampling in some strata (Stehman & Selkowitz, 2010). The Alaska NLCD accuracy assessment sample was selected from 30 of the 31 geographic strata and the results therefore represent these 30 conterminous strata. The western Alaska Peninsula was excluded because of the substantial additional costs and elevated risks of flying to this extremely remote region.

We partitioned the state of Alaska into 27-km × 27-km grid cells ("blocks"), with this block size chosen on the basis that all sample pixels within the block could be visited in one day of flight time using a fixed-wing aircraft. Before selecting the sample, blocks with greater than 25% of their land area located in adjacent Canada, and blocks containing less than 50% combined area of land and inland water were excluded. In total, 96.5% of the area of Alaska was eligible to be sampled after excluding these blocks and excluding the western Alaska Peninsula geographic stratum. Each remaining block was assigned to the geographic stratum covering the largest area within the block. A single block was randomly selected from each stratum with all blocks within a stratum having an equal probability of being selected (Fig. 1, inset).

Each 27-km × 27-km sampling block consisted of 900 rows by 900 columns of 30-m pixels. Nine east–west flight lines were located by randomly selecting a starting row, r_0, between 1 and 900, and then moving a sampling interval of 100 rows in both directions to complete the sample of 9 lines (rows) within a block. Each flight line was assigned a random number, and flight lines were sampled sequentially from the