Environ Monit Assess (2017) 189:19
DOI 10.1007/s10661-016-5745-x

Area-based fuzzy membership forest cover comparison between MODIS NPP and Forest Inventory and Analysis (FIA) across eastern U.S. forest

Youngsang Kwon · Bradley Wayne Baker

Received: 19 June 2016 / Accepted: 6 December 2016

Abstract This research assessed the accuracy of the moderate resolution imaging spectroradiometer's (MODIS) land cover classification of softwood and hardwood using a fuzzy-based approach for 31 easternmost states in the U.S. Our main objective was to quantitatively evaluate spatially explicit land cover classifications of MODIS net primary product (NPP) scheme using the USDA Forest Service's (FS) field-based, tree-specific Forest Inventory Analysis (FIA). We used a grid of 648 km^2 hexagons as base mapping units and interpreted our results at the USDA FS level IV ecological regions. Forest area was calculated for both MODIS and FIA and were found to be strongly correlated (Pearson's $r = 0.875$, $p < 0.01$), which suggests the two classifications are comparable. Area-based fuzzy memberships of softwood and hardwood forest were determined for both MODIS and FIA for each hexagon. We used cross-entropy (H_c) to evaluate the accuracy of the MODIS classification. Our results determined that the accuracy of MODIS forest cover classification was not uniform for all ecological regions. Tree species importance values (IV) and Shannon's diversity index (H_s) were calculated to examine species abundance and heterogeneity, which may partially explain discrepancies between MODIS and FIA classifications. The greatest misclassifications were due to (1) MODIS underestimating softwood forest cover and (2) MODIS confusing forest cover with other land covers such as grassland, cropland, or woody savanna. Our results provide a guideline for users to understand the degree of uncertainty of MODIS forest cover classifications in the eastern USA.

Keywords Fuzzy classification · MODIS · FIA · Field data · Cross-entropy

Introduction

The continuous monitoring of carbon cycle across Earth's entire vegetated land surface is a primary goal of Terra MODIS (or moderate resolution imaging spectroradiometer). With the on-going climate change, to better understand forest dynamics and its carbon cycle is important for managing timberland for the forest industry (Keenan 2015), balancing the carbon budget (Schaphoff et al. 2016), exploring the human-environment interactions (Canadell et al. 2007), and reacting to future climate change (Nemani et al. 2003; Reyer et al. 2015).

MODIS provides a wide ranging standard suite of global land products including reflectance (MOD09), temperature/emissivity (MOD11), land cover

Electronic supplementary material The online version of this article (doi:10.1007/s10661-016-5745-x) contains supplementary material, which is available to authorized users.

Y. Kwon (✉) · B. W. Baker
Department of Earth Sciences, University of Memphis, 1 Johnson Hall, Memphis, TN 38152, USA
e-mail: ykwon@memphis.edu

B. W. Baker
e-mail: bwbaker1@memphis.edu

Published online: 15 December 2016

classification (MOD12), vegetation index (MOD13), leaf area index/fraction of photosynthetic active radiation (MOD15), and derived modeled products of gross primary production/net primary production (GPP/NPP) (MOD17). Among the series of MODIS products, MOD12 data play a critical role in modeling other subsequent MODIS products. MOD17 algorithm, for example, is based on the radiation use efficiency logic (Monteith 1972) which relies heavily on the biome-specific physiological efficiency conversion parameter of ε based on MOD12 classification. The MOD12 products of forest land covers have become especially important given the pace and extent of human-induced deforestation and forest degradation. Any misclassification of land covers are regarded as a major factor resulting in high uncertainty of biomass/carbon budget estimation.

Although MODIS land cover validation at a global scale has been updated yearly with improvements, the efforts to evaluate regional scale classification in U.S. forest are very limited. At global scale, the accuracy of the International Geosphere Biosphere Program (IGBP) classification—the primary land cover scheme—of MODIS (Collection 5, current version) is estimated to be 74.8% using cross-validation of a training site database (Olofsson et al. 2012; Stehman et al. 2012). At regional scale, however, most validation efforts have been based on finer spatial resolution imagery rather than field-based ground truth data due to the small number of standardized field data covering large geographic extent._ENREF_30 Hao and Gen-Suo (2012) assessed the spatial agreement between three satellite-derived land cover products including MODIS, global land cover 2000 (GLC2000), and the National Land Cover/Use Datasets (NLCD) over China from 1999 to 2000 and determined that MODIS had greater spatial disagreement over regions of heterogeneous vegetation. While MODIS provides sub-classes of forest cover—evergreen needleleaf, evergreen broadleaf, deciduous needleleaf, and deciduous broadleaf—forest within MODIS pixels are often consist of heterogeneous forest types without identifiable discrete boundaries.

MODIS land cover validation requires extensive in situ data collected from comprehensive research and monitoring programs to derive and interpret broad-scale environmental conditions. A promising field-based MODIS validation candidate spatially covering conterminous USA and temporally consistent on a yearly basis is the Forest Inventory and Analysis (FIA) data managed by the USDA Forest Service. The FIA Program adopted a standardized nationwide annual inventory methodology in 1998 that enables spatially unbiased and timely monitoring of forest conditions (Bechtold 2005). This nationwide FIA annual inventory has greatly improved upon the less frequent state by state periodic inventory, which used a variety of survey standards that may have created biased geographic variations across states.

The use of field-based FIA plot data is valuable because it is completely independent from MODIS, and its numerous arrays of ground plot network allow the validation of various MODIS products across many different plant community types and biophysical conditions (Muukkonen and Heiskanen 2007). However, the desired comparability between coarse resolution of MODIS pixels (1 km by 1 km) and plot-level FIA has been problematic due to inherent scaling mismatches. The mismatch is that, while MODIS products have its spatial resolution of 500 m to 1 km, FIA data has spatial sampling intensity of one plot for every 24.3 km^2, with that plot itself having a total area of 0.0041 km^2 (McRoberts et al. 2005). A widely used technique to overcome spatial mismatches is to bridge spatial gaps using intermediate spatial resolution data such as Landsat satellite (Townshend and Justice 2002). However, this bridging approach propagates uncertainty in the multi-step scaling up process, and the spatial coverage is often limited to only a few of the many plant community types present in the USA. With the increasing needs of FIA combined with various remote sensing imageries, Riemann et al. (2010) provided a comprehensive assessment protocol for continuous variables such as biomass estimates derived from FIA, but the categorical land cover classification is not critically examined where aggregation-related biases are created and propagated in remote sensing data.

Kwon and Larsen (2012, 2013) demonstrated the use of plot-level FIA data to validate pixel-level MODIS-derived gross primary productivity (GPP) estimates by applying a series of screening variables of forest conditions recorded in FIA database (FIADB). Although not spatially explicit, a higher agreement of land cover classification between MODIS and FIA resulted in better MODIS GPP prediction by

FIA tree growth data. They classified species into taxonomic subdivision-level class forest types (i.e., gymnosperm and angiosperm, hereafter softwood and hardwood, respectively) by using cut off values of 75%. For example, each FIA plot was classified as hardwood or softwood if more than 75% of both its basal area and number of stems were hardwood or softwood, and plots containing <75% of one type were classified as mixed forest. This hard classified method, although commonly applied to both MODIS and FIA, created a mid-latitude dominant mixed forest type, which made the classification accuracy difficult to interpret at the level of forest types.

Fuzzy classification (or sub-pixel soft classification), on the other hand, is fuzzy membership based approach often used in mixed-class areas. Examples include suburban land cover classification (Zhang and Foody 1998), grassland classification (Sha et al. 2008), or the study of long-term vegetation changes (Okeke and Karnieli 2006). Fuzzy classification uses continuous zones of intermediate classes between the end-member classes rather than imposing arbitrary cut off values to delineate boundaries of the end-member classes (Hill et al. 2007). We adopted fuzzy membership concept evaluated by cross-entropy values to overcome scaling mismatches between MODIS and FIA.

The primary objective of this study is to evaluate spatially explicit land cover classification of MODIS (MOD12) using plot-based, tree-specific FIA data. First, we develop area-based fuzzy membership to derive spatially explicit forest types of softwood and hardwood for both MODIS and FIA. Second, we provide regional scale assessment of land cover classification between MODIS and FIA using cross-entropy values calculated at level-IV ecological regions and present descriptive results at level-III ecological regions. Finally, we discuss the potential cause of forest cover mismatches between two data sets.

Materials

Mapping framework

The study area is the 31 easternmost U.S. states. We use a grid of 648 km^2 hexagons for the basic mapping unit which had been used as the basis for the FIA sampling design (McRoberts et al. 2005) to maintain nationwide original FIA sampling intensity. The hexagonal grid framework, originally used in the Forest Health Monitoring (FHM) program, systematically aggregates plots and pixels independently from potentially regular spaced landscape features (White et al. 1992). We use ecological regions defined by the USDA FS for the interpretation unit. Ecological regions aggregate areas with similar characteristics into homogeneous regions based on temperature, precipitation, vegetation, natural land covers, and terrain features._ENREF_21 ecological regions are subdivided into four hierarchical levels, where each successive level is comprised of smaller nested regions. The study area is bounded by 12 level-III provinces and 66 level-IV sections, which we used to map spatial patterns that are difficult to represent at the hexagon mapping unit. A U.S. ecological regions shapefile was downloaded from the USDA FS (http://www.fs.fed.us/rm/ecoregions/products/map-ecoregions-united-states/). Hexagons are aggregated to the closest level-IV ecological region, and fuzzy memberships and cross-entropy values are averaged for each ecological region.

FIA database

FIA program uses systematic five-year rolling annual inventory system with the unified plot design. FIA plot design consist of four 7.2 m fixed-radius subplots to tally all trees with a diameter at breast height (d.b.h) of at least 12.7 cm and each subplot contains a microplot for seedlings and understory inventory (Bechtold 2005). The locational accuracy issues mandated by public law allowed perturbed latitude and longitude of plot locations, and a swapped small percentage of plots located on private lands with another similar condition of private plot (McRoberts et al. 2005; Woodall et al. 2009). However, these issues will be negligible in this study since plots are aggregated to large geographic extents, while these location perturbations are randomly applied within a 0.8 km radius of the actual location. FIA program manages tree census information such as the species, size, and health of trees under the relational database management system (RDBMS) to estimate status and trends in forest area. The database is available to the public via both Microsoft Access and plain text file format. FIA MS Access version of database (FIADB, version 6.0) was downloaded for the 31 easternmost U.S. states from the FIA DataMart (http://apps.fs.fed.us/fiadb-downloads/datamart.html). We used the most recent complete five-year inventory cycle as most states inventoried from year 2009 to 2013.

MODIS NPP land cover (MOD12)

The MOD12 product supplies five classification schemes at annual time steps and 500-m spatial resolution for 2001 to present. The primary land cover scheme is provided by an IGBP land cover classification (Friedl et al. 2010), and there are other classification schemes including the University of Maryland classification scheme (Hansen et al. 2000), the MODIS-derived net primary production (NPP) classification scheme described by Running et al. (1994), the MODIS-derived LAI/fPAR Biome scheme described by Myneni et al. (1997), and the plant functional type (PFT) scheme described by Bonan et al. (2002). We focus MODIS-derived net primary production (NPP) classification scheme because this scheme does not use a mixed forest type thus enables all species classified by either hardwood or softwood. The NPP classification scheme comprises eight land cover types including water, evergreen needleleaf, evergreen broadleaf, deciduous needleleaf, deciduous broadleaf, annual broadleaf, annual grass, non-vegetated land, and urban. NPP forest pixels are reclassified as either hardwood or softwood, and this allows a direct comparison to FIA taxonomic subdivision-level tree classes through a fuzzy membership agreement approach. MODIS database (MOD12Q1, collection 5) was downloaded for the 31 easternmost U.S. states from the USGS Land Processes Distributed Active Archive Center Data Pool (https://lpdaac.usgs.gov/data_access/data_pool). We used the year 2010 because this represented the intermediate year of the FIA five-year inventory cycle.

Methods

Matching forest categories between MODIS and FIA

We reclassified both MODIS forest related land cover classes and FIA tree-level species code to softwood and hardwood to match forest categories between MODIS and FIA. MODIS forest type land covers are reclassified from evergreen needleleaf and deciduous needleleaf forest to softwoods, evergreen broadleaf and deciduous broadleaf forest to hardwoods, respectively. Total 254 tree species identified for the eastern U.S. forest were also manually grouped to either softwoods or hardwoods using species group codes assigned to each tree species. FIA database contains 54 species group codes classified by eastern and western regions of softwood and hardwood for reporting purposes.

Forest area calculation of FIA

We first calculated forest area by utilizing the plot-specific area expansion factor (variable code EXPCURR) and its adjustment factor (variable code ADJ_EXPCURR) stored in PLOTSNAP table of FIADB. The EXPCURR variable represents plot-specific sampling intensity which can be interpreted as the number of acres each plot represents. The area adjustment factor (ADJ_EXPCURR) is used to compensate for the proportion of plots not sampled due to denied access or inaccessible locations. These areas may differ each time new plots replaces older plots under the annual inventory system. To calculate forest area, therefore, EXPCURR is multiplied by ADJ_EXPCURR to estimate the forested area a plot represents. Under the annual inventory system, most states follow the EXPCURR value of approximately 6000 representing the sampling intensity of one plot identified for approximately 6000 acres (~24.3 km^2) of forest area, which is based on the national precision of 3% per million acres in Eastern U. S. (Bechtold 2005). There are five states—Minnesota, Wisconsin, Delaware, Indiana, and Rhode Island—that adopted a double sampling intensity, so their EXPCURR is approximately half of other regular states thus exceeds national standard precision. The estimated area of each plot is then mapped as a uniform circular shape centered at corresponding plot coordinates. Finally, we calculated a percent of forest area at each hexagon by spatially intersecting circular area by hexagonal grid.

Fuzzy membership of FIA forest type

Fuzzy membership of softwood and hardwood is constructed based on relative abundance of each species summed by softwood and hardwood within 648 km^2 of hexagon. To calculate relative abundance of species at hexagonal mapping units, we first used per-acre basis tree expansion factor (variable code TPA) stored in TREE table to standardize actual area of measured plots. Different plot area is possible when a plot includes individual subplots where no accessible area is present. Under the current fixed-plot designs, the TPA value is represented by one tree equal to the inverse of the plot area in acres as:

$$\mathrm{TPA} = \frac{1}{N*A} \tag{1}$$

where N is the number of subplots, and A is the area of each subplot.

Then, we calculated species-level importance values (IV) at each hexagon as follows:

$$\mathrm{IV(x)} = 0.5*\sum(\mathrm{TPA}_i*\mathrm{BA}_i(\mathrm{x}))/\sum(\mathrm{TPA}_i*\mathrm{BA}_i(\text{all species})) + 0.5*\sum(\mathrm{TPA}_i*\mathrm{NS}_i(\mathrm{x}))/\sum(\mathrm{TPA}_i*\mathrm{NS}_i(\text{all species})) \tag{2}$$

where IV(x) is importance value of species x calculated at a hexagon, TPAi is tree per acre at plot i, x is a particular species in plot i, BA is basal area, and NS is number of stems.

Given in Eq. 2, summation of softwood IV and hardwood IV results in exclusive binomial membership as summation of all species IV equals to one at a hexagonal grid. Forest area of softwood and hardwood within a hexagon is then estimated as forest area calculated from above section (forest area calculations of FIA) multiplied by membership of each softwood and hardwood of IV. Finally, the area-based fuzzy membership of softwood and hardwood is calculated by forest area of softwood and hardwood within each hexagon divided by the area of hexagonal grid (648 km^2). The FIA area-based membership of each forest type ranges from 0 to 1.

Fuzzy membership of MODIS forest type

We first determine the total count of MODIS softwood and hardwood pixels within each 648 km^2 hexagonal grid. MOD12 pixels have a resampled spatial resolution of 500 m resulting in each individual pixel area of 0.25 km^2. The total count of softwood and hardwood pixels is then multiplied by the above pixel area to determine the softwood and hardwood forest area of each hexagon, respectively. The total MODIS forest area within each hexagon is calculated by summing the softwood and hardwood areas. Area-based fuzzy membership of each forest type is then simply calculated by MODIS actual area of softwood and hardwood within each hexagon divided by the area of 648 km^2 hexagonal grid. The MODIS area-based membership of each forest type ranges from 0 to 1.

Fuzzy membership assessment

We first apply Pearson's r correlation to FIA and MODIS forest areas to determine if the two are comparable. Then, we evaluate the accuracy of area-based softwood and hardwood fuzzy membership by applying an equation that measures cross-entropy (H_c) between FIA and MODIS calculated as (Foody 1995):

$$H_c = -\sum \mathrm{p(x)}\log_2 \mathrm{p}'(\mathrm{x}) + \sum \mathrm{p(x)}\log_2\ \mathrm{p(x)} \tag{3}$$

where p is the MODIS forest type classification and p is the FIA forest type classification.

A cross-entropy (H_c) value of zero indicates perfect agreement between the two classifications, while increasing positive and negative values indicate disagreement. Therefore, greater disagreement occurs when values are further away from zero.

Forest heterogeneity and classification accuracy

To examine the overall influence of species heterogeneity to the MODIS classification accuracy, Shannon's diversity index (H_s) is calculated at each hexagonal grid. The H_s value is commonly used to characterize species diversity in a community as it accounts for both abundance and evenness of the species present and is calculated as (Begon et al. 1996):

$$H_s = -\sum p_i \ln p_i \quad (4)$$

where H_s is Shannon's diversity index, p_i is the proportion of species i relative to the total number of species, and ln is the natural logarithmic function.

The overall relationship between forest heterogeneity calculated by H_s and classification accuracy by cross-entropy (H_c) value at each hexagonal grid is then examined by Pearson's r value.

Results

This study compares areas of softwood and hardwood land covers at a grid of 648 km^2 hexagons classified by MODIS and FIA to evaluate the agreement between two data sets. Field-based FIA plot data is used as ground truth reference to verify the MODIS forest classifications. Fuzzy memberships of forest area classified as softwood and hardwood are calculated for both FIA and MODIS. Cross-entropy analysis is implemented to evaluate the agreement between both classifications. We first describe FIA and MODIS forest areas and fuzzy memberships at the hexagonal mapping unit. Then, to better highlight spatial patterns, we report FIA and MODIS forest type classifications aggregated to 66 level-IV ecological regions (Supplementary material, Table 1). We describe the spatial patterns based on 12 level-III ecological regions (Fig. 1).

Fuzzy membership of forest area

FIA and MODIS forest area shows similar geographic patterns (Fig. 2), and the forest areas at hexagonal grids between two data sets are strongly correlated (Pearson's r = 0.875, p < 0.01). The strong spatial agreement between MODIS and FIA forest area confirms the compatibility of MODIS and FIA data. This also reflects the improved national consistency of FIA, which was achieved through a national sampling design and plot configuration adopted in annual FIA since 1998. The largest forest areas for both FIA and MODIS are located in the northeast, Upper-Midwest, along the Appalachian Mountains, and in the southeast region (Fig. 2). The smallest forest areas are the continental interior, Mid-Atlantic, southern Florida, and the Lower Mississippi River Valley for both classifications (Fig. 2). There are a few hexagons with substantial differences in forest area between FIA and MODIS. Hexagons with greater FIA forest area (Fig. 2a) than MODIS are mostly located in the continental interior. Higher percentages of MODIS forest area (Fig. 2b) than FIA are found in the northeast, Upper-Midwest, and in the southern and central Appalachians. Fuzzy memberships of softwood and hardwood of both FIA and MODIS show similar spatial patterns at the hexagonal grid (Figs. 3 and 4). While softwood forest types of FIA and MODIS are mostly confined to the southeast, and smaller areas are found in the northeast and Upper-Midwest for both classifications (Fig. 3); FIA has greater percentages and a larger spatial footprint of softwood forest in all of these areas. MODIS fuzzy membership for hardwood shows overall higher percentage than FIA with greater magnitude in the southeast, Florida, northeast, along the Appalachian Mountains, and in the Upper-Midwest (Fig. 4). FIA fuzzy membership shows overall scattered low to intermediate hardwood percentages (15 to 71%) throughout study area except in continental interior region (greater than 72%) (Fig. 4).

FIA classification by ecological region

FIA shows lower percentages of softwood forest area in the USA compared to hardwood. A latitudinal gradient is shown, with higher softwood percentages located at southern and northern latitudes, and lower softwood percentages in the middle latitudes (Fig. 5a). Regions in the southern latitudes that have the highest percentage of softwood forest area (26 to 49%) include the Southeastern Mixed Forest, Ouachita Mixed Forest, and Outer Coastal Plain. In the northern latitudes, the Laurentian Mixed Forest and Adirondack regions are classified as 26 to 57% softwood. The continental interior in the middle latitudes has the least amount of forest area classified as softwood. The Eastern Broadleaf (Continental), Eastern Broadleaf (Oceanic), Prairie Parkland, and Eastern Broadleaf (Oceanic) in the Mid-Atlantic are classified as less than 8% softwood. The FIA data shows a northeast to southwest

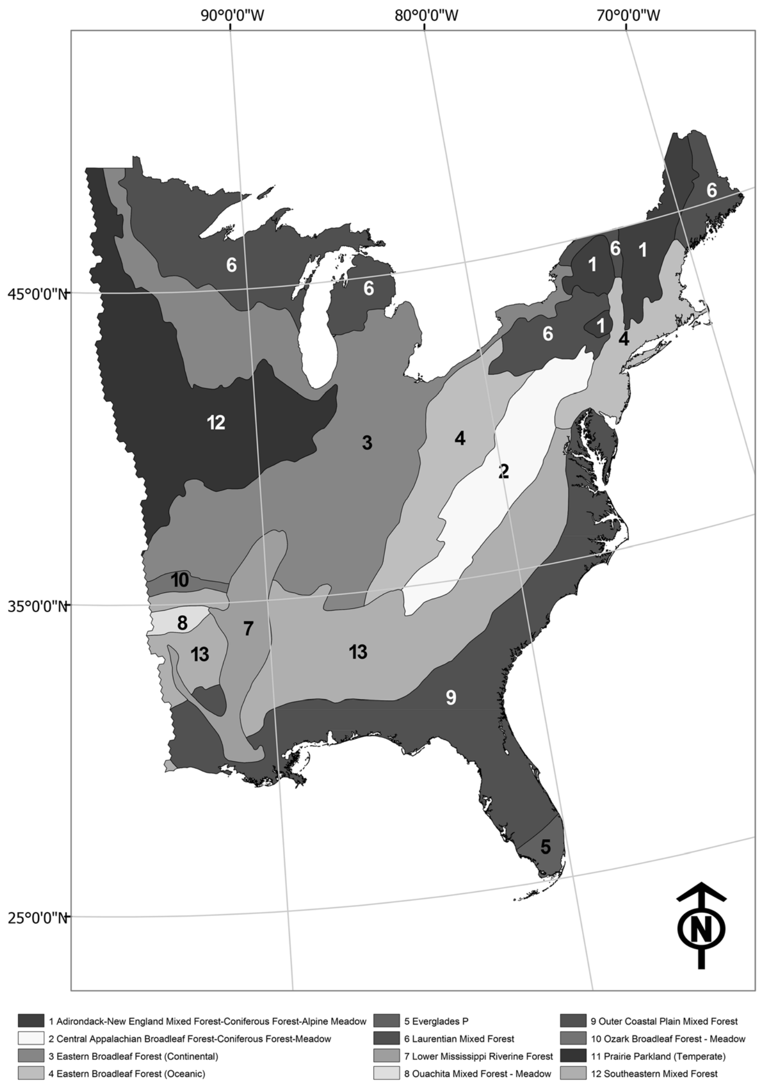

Fig. 1 Twelve EPA level-II ecological regions covering 31 easternmost states

gradient of hardwood forest along the Appalachian Mountains (Fig. 6a). Between 43 and 99% of the forest area is classified as hardwood for both the Eastern Broadleaf (Oceanic), Central Appalachian, Adirondack, and Eastern Broadleaf (Continental) regions. The Southeastern Mixed Forest, Eastern Broadleaf (Continental), Ozark Mixed Forest, and Laurentian Mixed Forest are classified as 29 to 56% hardwood. The regions with the least amount of forest area classified as hardwood are along coastal regions and in the Midwest. The Outer Coastal Plain, Everglades, Prairie Parkland, and Eastern Broadleaf (Continental) in the Midwest have the lowest percentages (less than 28%) of forest area classified as hardwood.

MODIS classification by ecological region

Most of the forest area classified as softwood by MODIS is confined to regions in the southeast, northeast, and Upper-Midwest (Fig. 5b). This pattern is similar to the latitudinal gradient shown by the FIA softwood classification, but the spatial extent of the MODIS classification is not as extensive. The maximum percentages of MODIS softwood (42 to 49%) occur in the Laurentian Mixed Forest in the Upper-Midwest. The Laurentian Mixed Forest has the highest softwood percentages in the northern latitudes between 17 and 49%. The highest percentage of softwood forest in the southern latitudes occurs in the Southeastern Mixed Forest and the

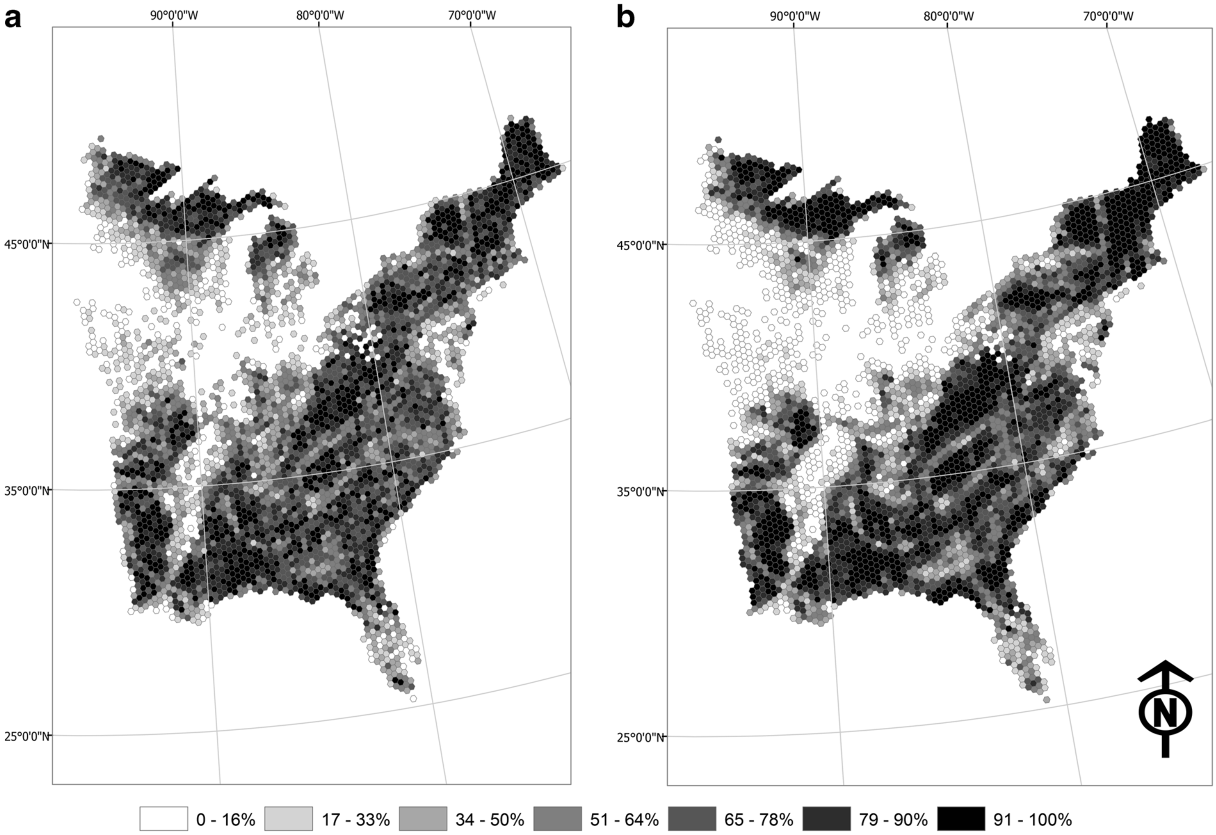

Fig. 2 Percent of forest area of each 648 km^2 hexagon classified by **a** FIA and **b** MODIS. Class breaks determined using seven equal intervals

Outer Coastal Plain (17 to 33%). The lowest softwood percentages (<8%) are found in the Eastern Broadleaf (Oceanic), Eastern Broadleaf (Continental), Prairie Parkland, Adirondack, Lower Mississippi Riverine Forest, Outer Coastal Plain, and Ouachita Mixed Forest. Similar to the FIA hardwood forest area, MODIS hardwood classifications shows a northeast to southwest gradient (Fig. 6b). One exception to this gradient is the Lower Mississippi Riverine Forest, which has a lower percentage (29 to 42%) of forest area classified as hardwood. The Central Appalachian, Adirondack, and Laurentian regions have the highest percentages of hardwood forests with 86 to 99%. Along the same linear gradient, the Southeastern Mixed Forest and Eastern Broadleaf (Continental), Ozark Broadleaf Forest, Ouachita Mixed Forest regions are classified as 57 to 85% hardwood forest. The continental interior has the lowest percentages (less than 28%) of hardwood forest area. These regions include Prairie Parkland and the Eastern Broadleaf (Continental).

Comparison between FIA and MODIS

An examination of FIA and MODIS softwood forest area (Fig. 5) shows a similar spatial pattern between both classifications, but FIA classified more forest area as softwood compared to MODIS. A latitudinal gradient of higher percentages of softwood forest area is found at southern and northern latitudes. These are separated by lower softwood percentages in the middle latitudes. Although the general softwood spatial pattern is similar between the two classifications, MODIS is more confined to the Outer Coastal Plain and Southeastern Mixed Forest in the southern latitudes. Higher softwood

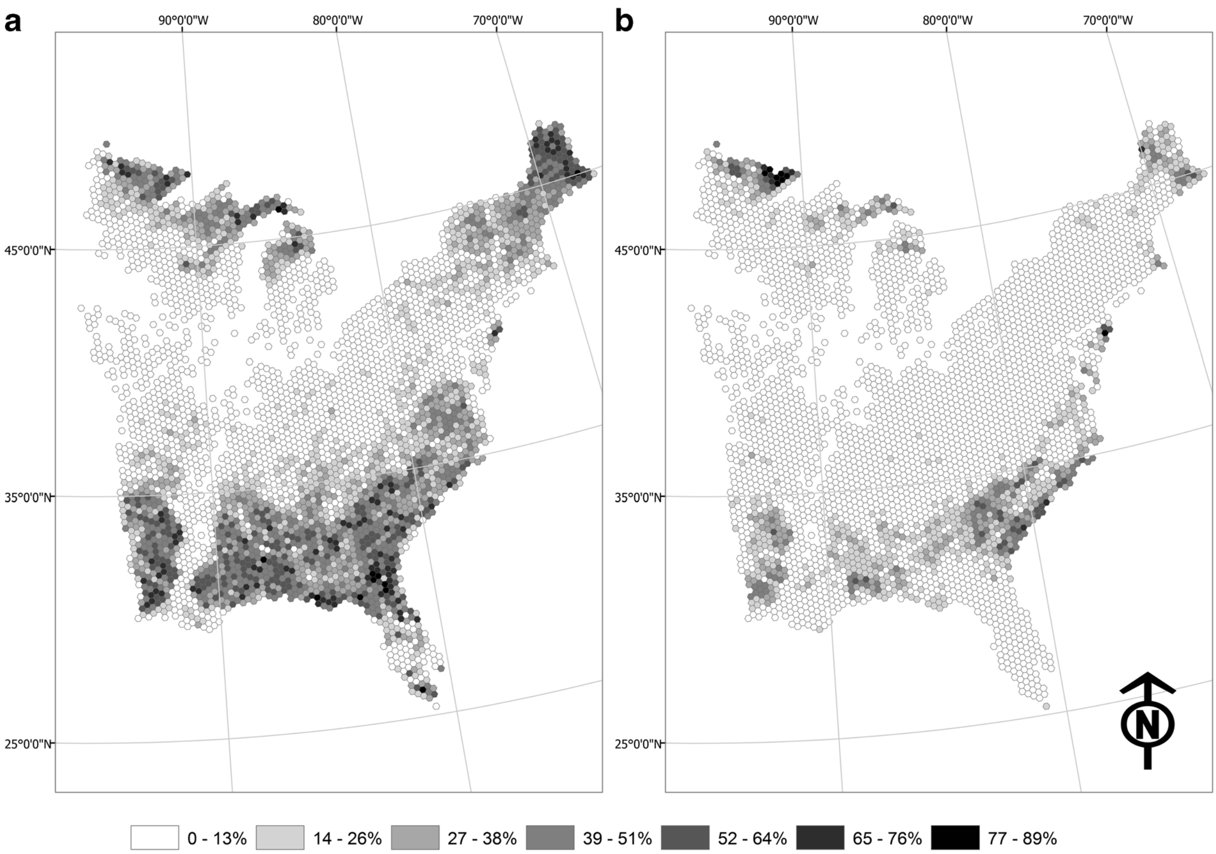

Fig. 3 Fuzzy membership of softwood forest (%) of each 648 km^2 hexagon classified by **a** FIA and **b** MODIS. Class breaks determined using seven equal intervals

percentages classified by both FIA and MODIS are found in the Northeast and Upper-Midwest. FIA has higher softwood percentages compared to MODIS in the Southeastern Mixed Forest, Ozark Broadleaf, Outer Coastal Plain, Ouachita Mixed Forest, Laurentian Mixed Forest, Eastern Broadleaf (Oceanic), Everglades, Central Appalachian Mixed Forest, and Adirondack regions.

A comparison of FIA and MODIS hardwood forest classifications (Fig. 6) shows similar spatial patterns, but there are noticeable differences. A northeast to southwest gradient of maximum percentages of hardwood forest area is indicated by both classifications. The gradient follows along the Appalachian Mountains into the lower Mississippi River Valley. The lowest percentages of hardwood forest are located in the continental interior for both classifications. A major difference between FIA and MODIS hardwood forest is in the Adirondack and Laurentian regions in the Northeast, Outer Coastal Plain Mixed Forest, and the Southeastern Mixed Forest, where MODIS has higher hardwood percentages. FIA has higher percentages of hardwood forest in both Easter Broadleaf (Continental) and Prairie Parkland regions.

Fuzzy membership assessment

Cross-entropy values (H_c) are calculated to determine agreement or disagreement between FIA and MODIS. H_c values near zero suggest that FIA and MODIS are in close agreement, and increasing positive and negative values suggest disagreement. Positive H_c values mean that MODIS classified a

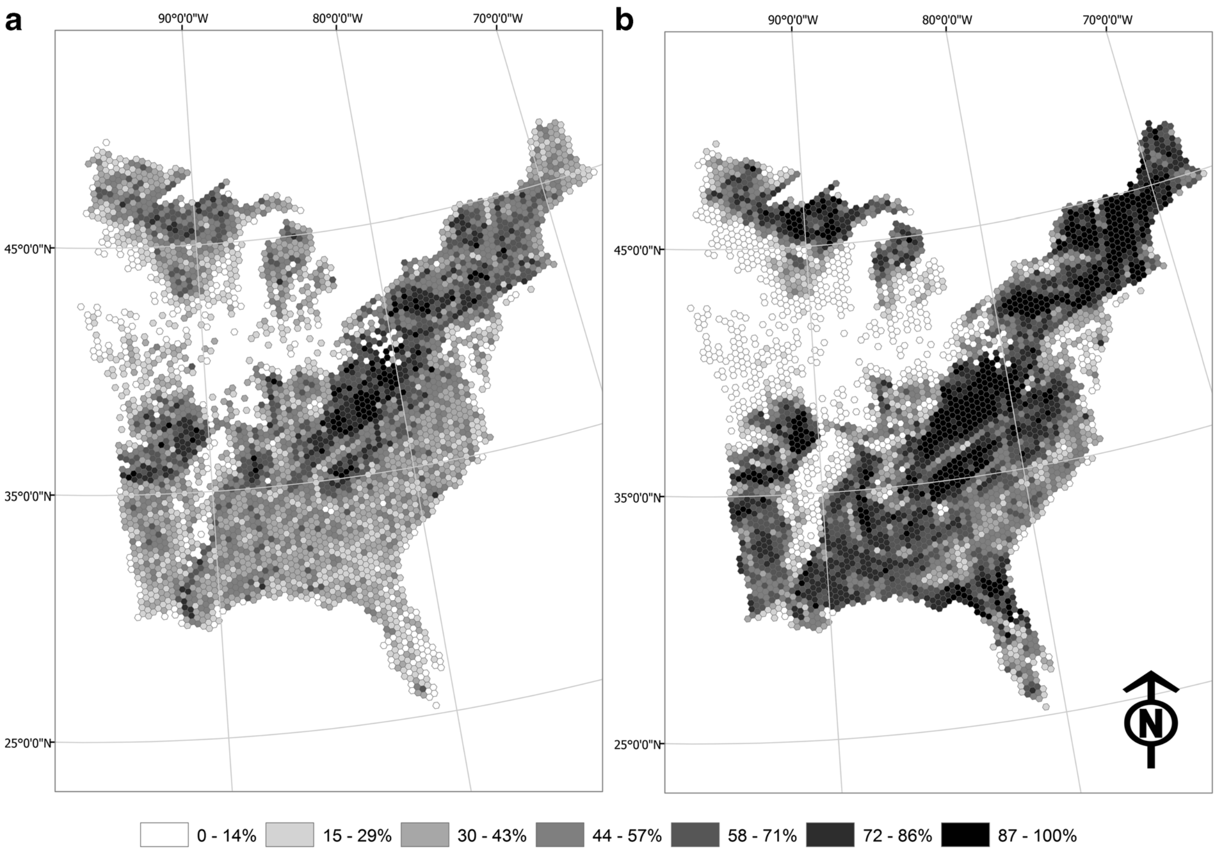

Fig. 4 Fuzzy membership of hardwood forest (%) of each 648 km^2 hexagon classified by **a** FIA and **b** MODIS. Class breaks determined using seven equal intervals

higher percentage of one forest type compared to FIA. Negative H_c values mean that FIA classified a higher percentage of one forest type compared to MODIS. For example, the Eastern Broadleaf Forest (Continental) has negative H_c values (−0.13 to 0.00) because FIA classified the region with higher percentages of hardwood forest (14 to 42%) compared to MODIS (0 to 28%).

An examination of H_c values (Fig. 7) determined the highest disagreement occurs in the Southeast, Northeast, and along the Gulf Coast. The Adirondacks, Laurentian Mixed Forest, Everglades, Outer Coastal Plain, Ouachita Mixed Forest, and Southeastern Mixed Forest regions have the largest values (0.43 to 0.84), and therefore, the greatest disagreement. Large H_c values (0.15 to 0.42) are found in the Laurentian Mixed Forest, Eastern Broadleaf (Continental) in the southern Appalachians, and Eastern Broadleaf (Oceanic) forests in the northeast. Better agreement between FIA and MODIS is found in the continental interior, central Appalachians, and in the Mid-Atlantic. The Central Appalachian, Lower Mississippi Riverine Valley, Eastern Broadleaf (Oceanic), and Ozark Mixed Forest each have H_c values between 0.01 and 0.14, and negative H_c values (−0.13 to 0.00) are found in Prairie Parkland and Eastern Broadleaf (Continental), and Eastern Broadleaf (Oceanic) regions.

Heterogeneity to classification accuracy

We examined overall relationship between forest heterogeneity calculated by Shannon's diversity index (H_s) and classification accuracy by cross-entropy (H_c) value using Pearson's r value. The average H_s value of all hexagonal grids was 2.29 with standard deviation of

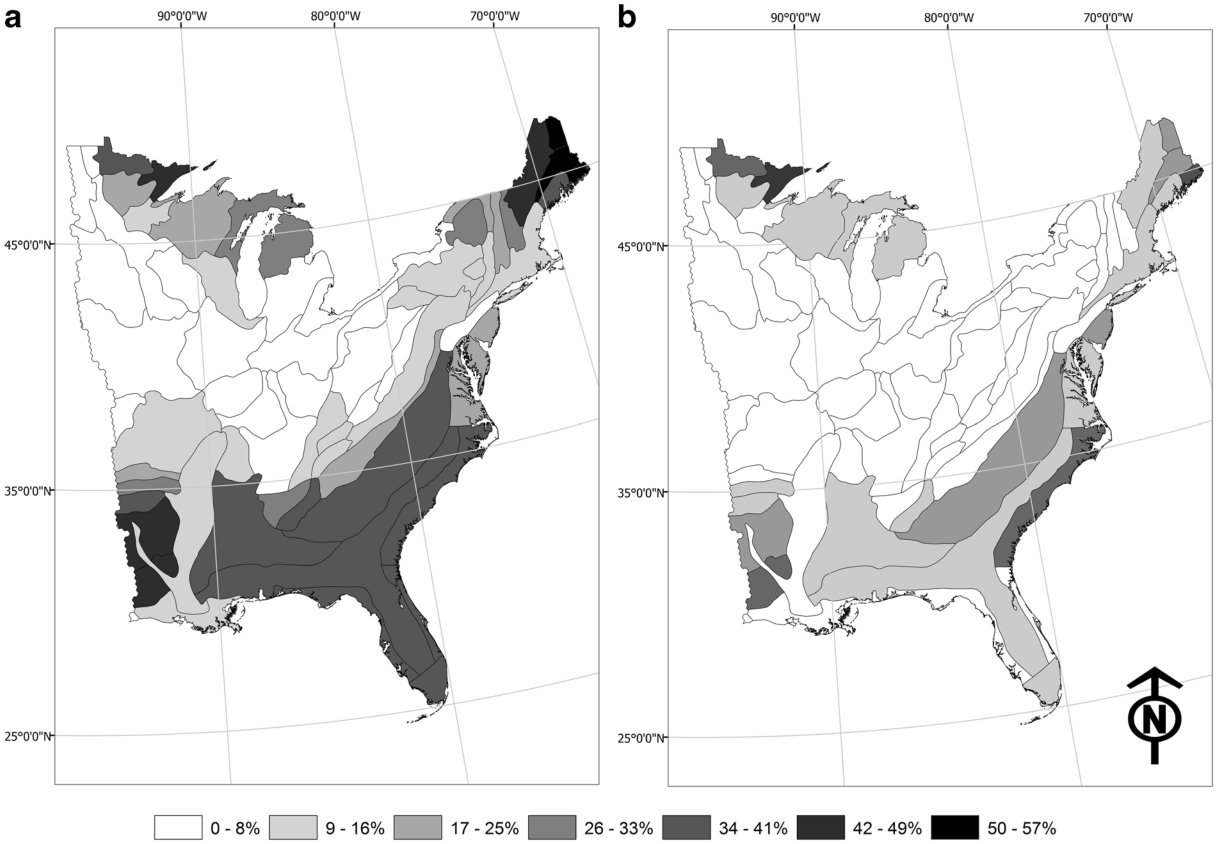

Fig. 5 Fuzzy membership of softwood forest (%) classified by **a** FIA and **b** MODIS averaged to level-IV ecological regions. Class breaks determined using seven equal intervals

0.39, and the average H_c value was 0.14 with standard deviation of 0.32. The relationship between the diversity (H_s) and the cross-entropy (H_c) was positively correlated (Pearson's $r = 0.21$, $p < 0.01$) meaning disagreement between FIA and MODIS increased with heterogeneity of species composition but the correlation was rather weak. This weak correlation may imply that while MODIS provides reliable taxonomic subdivision-level classification, its pixel resolution is not comparable to species level classification. In order to determine error patterns, more regional scale validation efforts are needed.

Discussion

To better discuss the agreement between MODIS and FIA forest classification, we examined species level IVs at each hexagonal grid and the MODIS IGBP scheme of land covers which include 17 total land cover classes. The reason we examined the IGBP scheme is because the MODIS NPP scheme does not include any non-forest or non-urban classes, thus it is difficult to determine confusion errors between forest and non-forest classes such as grasslands or croplands.

In terms of forest cover discrepancy between MODIS and FIA, when averaged by level-IV ecological regions, FIA estimated slightly higher (mean 2.03%) forest cover than MODIS with a confidence interval of ±0.21% ($p < 0.05$). The only ecological region beyond the confidence interval disagreement was Lake Agassiz, Aspen Parklands Section in Minnesota, where FIA estimated 23% higher forest cover than MODIS. FIA estimated total forest cover in that ecological

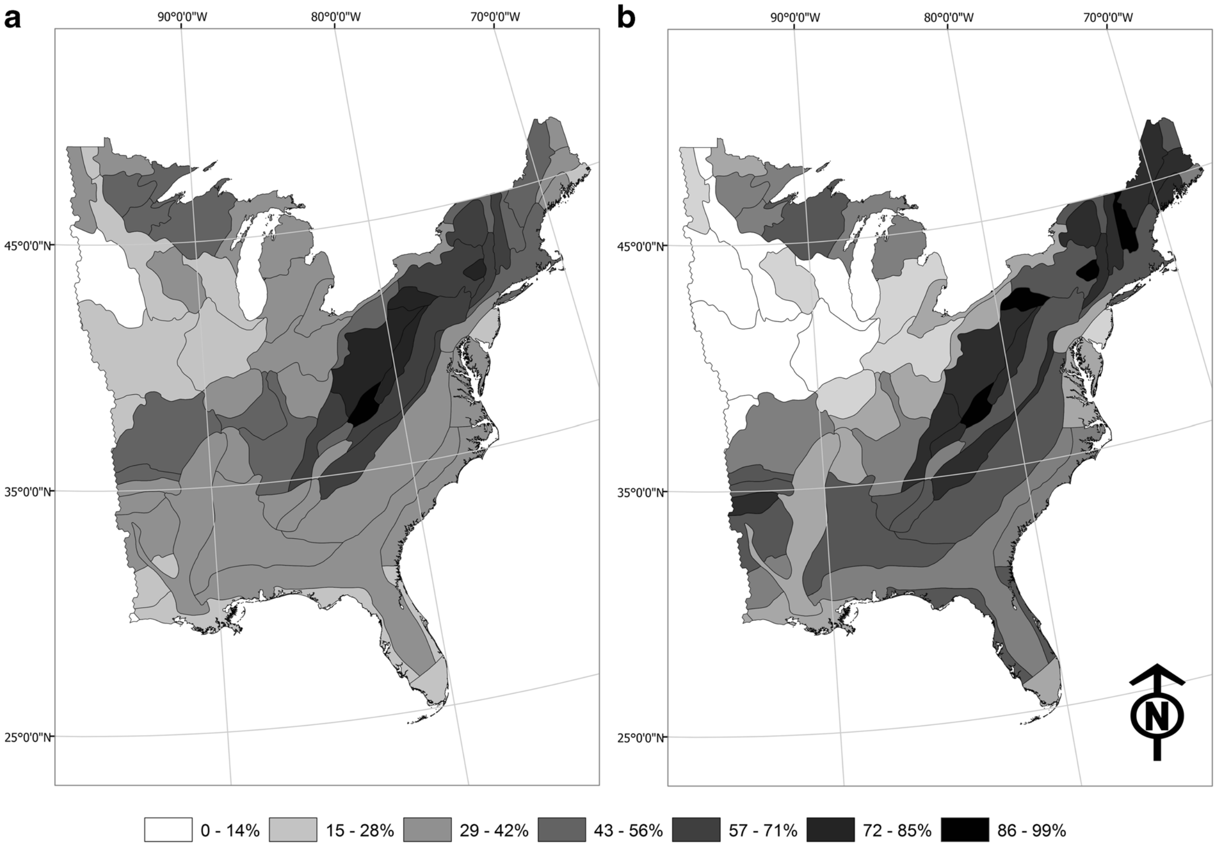

Fig. 6 Fuzzy membership of hardwood forest (%) classified by **a** FIA and **b** MODIS averaged to level-IV ecological regions. Class breaks determined using seven equal intervals

region as 345 km^2 while MODIS NPP scheme classified total forest cover as 43 km^2. The majority of tree species determined by IV values were Red Pine (IV 37.8), followed by Quaking Aspen (IV 26.8), and northern White-cedar (IV 19.1), thus this area is a good mixture of softwood and hardwood. MODIS IGBP scheme classified a majority of area as grassland (38%), followed by croplands (33%) and cropland/natural vegetation mosaic (15%). Within the forest cover class, the majority was classified as mixed forest (34 km^2) followed by the deciduous needleleaf forest (8 km^2). This area highlights the difficulty in distinguishing similar vegetative covers such as deciduous forest, croplands and grasslands in a coarse scale passive sensor (Wright and Wimberly 2013). In addition, land cover heterogeneity may add errors as this area is characterized as a complex mosaic of various small-patch land covers such as prairies, woodlands, forests on uplands, wet prairies, meadows, fens, and wet forests in wetlands (Omernik 2004).

In terms of fuzzy-classification discrepancy, the largest entropy value (greater than 0.5) was found in Florida Coastal Lowlands (both Eastern and Western section) and New England region (e.g., Central Maine Coastal & Interior, Aroostook Hills & Lowlands, and New England Piedmont section). Florida Coaster Lowlands exhibited very good forest cover agreement (FIA 0.65 and MODIS 0.67), but the forest type classification created large errors as FIA estimated an average 38% of softwood and 23% hardwood, whereas MODIS NPP estimated only 5% of softwood and 60% of hardwood. The forest composition in this area is dominated by Slash Pine (IV 16.23) and Loblolly Pine (IV 14.76), followed by Balsam Fir (IV 5.57) and Red maple (IV 5.03). On the other hand, MODIS IGBP classified the majority of the area as woody savannas (30.4%), followed by

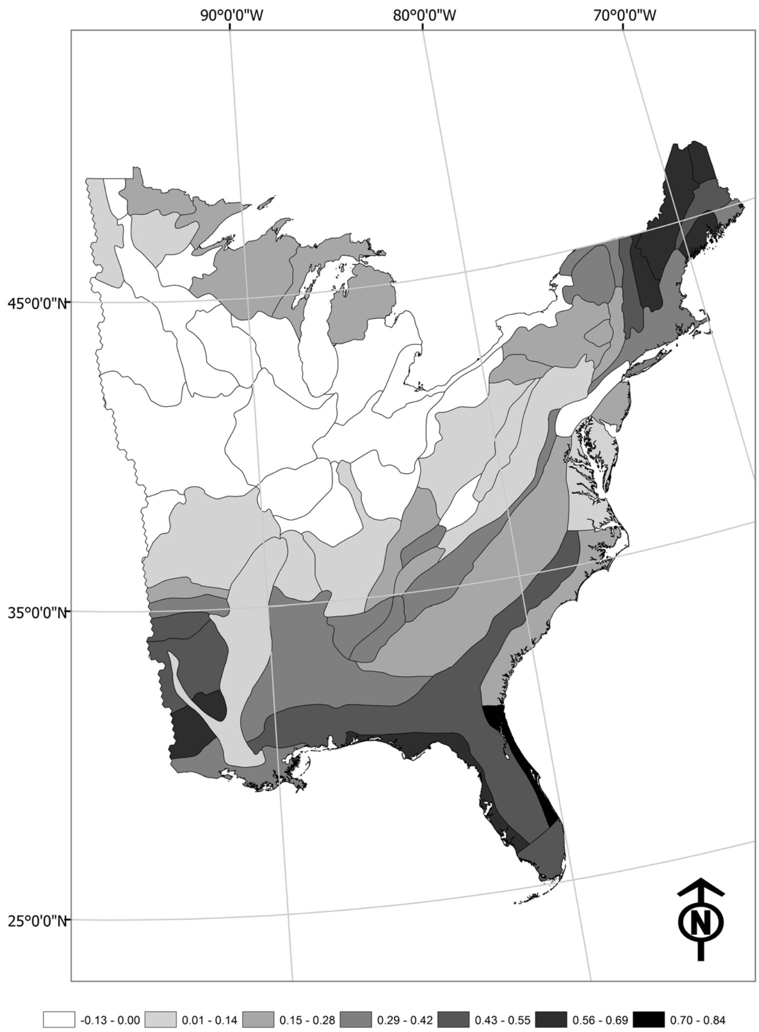

Fig. 7 Cross-entropy (H_c) values averaged to 66 level-IV ecological regions. Values closer to zero indicate better agreement between FIA and MODIS forest classifications, whereas values further from zero indicate greater disagreement. Class breaks determined using seven equal intervals

Evergreen broadleaf (20.6%) and Grassland (13.7%). MODIS NPP scheme failed to classify softwood in these coastal areas, and the error seemed to be the confusion between the woody savannas and softwood. The second largest disagreement found in the New England region is also related to missing classification of softwoods in MODIS NPP. The forest composition in the New England region is relatively evenly comprised. The most abundant tree species, Loblolly pine, has an IV of 14.08, followed by Red maple (IV 11.68) and Balsam Fir (IV 10.72). The MODIS IGBP scheme, on the other hand, classified the majority of area as mixed forest (65%), followed by 20% of deciduous broadleaf and only 3% of evergreen needleleaf. Similarly, MODIS NPP scheme classified it as hardwood (82%) dominant with less than 10% classified as softwood. Both Florida Coastal Lowland and New England regions can be characterized as a heterogeneous forest with high composition evenness, and it is not surprising that this heterogeneity deteriorates the accuracy in a remote sensing classification.

The MODIS carbon estimates (MOD17) are mainly validated by eddy flux towers such as Fluxnet or Ameriflux communities but they are mostly located in homogeneous managed forest plots. Thus, the accuracy of upstream input of land cover must be guaranteed to provide scientific credibility. As the carbon cycle models such as Biome-BGC (Running et al. 2004) or 3-PGS (Coops et al. 1998) widely utilize MODIS land covers to estimate continental scale GPP or NPP, this study can allow users to construct an upstream land cover error map and assess how the misclassified land cover classes affect prediction work.

Conclusions

This study aims to provide a "best practice" quantitative assessment of MODIS forest type classification validated by field-based FIA plot data. The result showed the accuracy of the MODIS NPP product is not uniform in all the regions assessed by applying a cross-entropy equation. We determined that while MODIS forest classifications are generally agreeable with field-based forest types of softwood and hardwood, further species group or species level classification is not comparable to field-based FIA plot data. Overall, MODIS exhibited a tendency to underestimate softwood forest at southern latitudes and in the northeast likely due to misclassification between forest cover and other vegetative land covers such as grassland, cropland, or woody savannas. The best forest classification agreement occurred over the continental interior, Central Appalachians and in the Mid-Atlantic. With the increasing need of accurate land cover inputs in modeling forest carbon dynamics, this study can provide a baseline reference for users to understand the extent and degree of uncertainty as well as a guideline for further regional scale validation of land cover classification.

References

Bechtold, W. A., & P., P. L. (Eds.) (2005). *The enhanced forest inventory and analysis program-national sampling design and estimation procedures*. Asheville: USDA Forest Service.

Begon, M., Harper, J., & Townsend, C. (1996). *Ecology: individuals, populations and communities* (3rd ed.). Oxford: Blackwell Science.

Bonan, G. B., Oleson, K. W., Vertenstein, M., Levis, S., Zeng, X., Dai, Y., et al. (2002). The land surface climatology of the community land model coupled to the NCAR community climate model. *Journal of Climate, 15*(22), 3123–3149. doi:10.1175/1520-0442(2002)015<3123:TLSCOT>2.0.CO;2.

Canadell, J. G., Le Quéré, C., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., et al. (2007). Contributions to accelerating atmospheric CO2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the National Academy of Sciences, 104*(47), 18866–18870.

Coops, N. C., Waring, R. H., & Landsberg, J. J. (1998). Assessing forest productivity in Australia and New Zealand using a physiologically-based model driven with averaged monthly weather data and satellite-derived estimates of canopy photosynthetic capacity. Forest Ecology and Management, 1041–3, 113–127, doi:10.1016/S0378-1127(97)00248-X.

Foody, G. M. (1995). Cross-entropy for the evaluation of the accuracy of a fuzzy land cover classification with fuzzy ground data. *ISPRS Journal of Photogrammetry and Remote Sensing, 50*(5), 2–12.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., et al. (2010). MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment, 114*(1), 168–182. doi:10.1016/j.rse.2009.08.016.

Hansen, M. C., Defries, R. S., Townshend, J. R. G., & Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing, 21*(6–7), 1331–1364. doi:10.1080/014311600210209.

Hao, G., & Gen-Suo, J. (2012). Spatial and quantitative comparison of satellite-derived land cover products over China. *Atmospheric and Oceanic Science Letters, 5*(5), 426–434.

Hill, R. A., Granica, K., Smith, G. M., & Schardt, M. (2007). Representation of an alpine treeline ecotone in SPOT 5 HRG data. *Remote Sensing of Environment, 110*(4), 458–467. doi:10.1016/j.rse.2006.11.031.

Keenan, R. J. (2015). Climate change impacts and adaptation in forest management: a review. *Annals of Forest Science, 72*(2), 145–167.

Kwon, Y., & Larsen, C. P. S. (2012). Use of pixel- and plot-scale screening variables to validate MODIS GPP predictions with Forest Inventory and Analysis NPP measures across the eastern USA. *International Journal of Remote Sensing, 33*(19), 6122–6148. doi:10.1080/01431161.2012.680615.

Kwon, Y., & Larsen, C. P. S. (2013). An assessment of the optimal scale for monitoring of MODIS and FIA NPP across the eastern USA. *Environmental Monitoring and Assessment, 185*(9), 7263–7277. doi:10.1007/s10661-013-3099-1.

McRoberts, R. E., Holden, G. R., Nelson, M. D., Liknes, G. C., & Gormanson, D. D. (2005). Using satellite imagery as ancillary data for increasing the precision of estimates for the Forest inventory and analysis program of the USDA Forest Service. *Canadian Journal of Forest Research, 35*(12), 2968–2980. doi:10.1139/x05-222.

Monteith, J. L. (1972). Solar radiation and productivity in tropical ecosystems. *Journal of Applied Ecology, 9*(3), 747–766. doi:10.2307/2401901.

Muukkonen, P., & Heiskanen, J. (2007). Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: a possibility to verify carbon inventories. *Remote Sensing of Environment, 107*(4), 617–624.

Myneni, R. B., Ramakrishna, R., Nemani, R., & Running, S. W. (1997). Estimation of global leaf area index and absorbed par using radiative transfer models. *IEEE Transactions on Geoscience and Remote Sensing, 35*(6), 1380–1393. doi:10.1109/36.649788.

Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., et al. (2003). Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science, 300*(5625), 1560–1563. doi:10.1126/science.1082750.

Okeke, F., & Karnieli, A. (2006). Methods for fuzzy classification and accuracy assessment of historical aerial photographs for vegetation change analyses. Part I: algorithm development. *International Journal of Remote Sensing, 27*(1), 153–176. doi:10.1080/01431160500166540.

Olofsson, P., Stehman, S. V., Woodcock, C. E., Sulla-Menashe, D., Sibley, A. M., Newell, J. D., et al. (2012). A global land-cover validation data set, part I: fundamental design principles. *International Journal of Remote Sensing, 33*(18), 5768–5788. doi:10.1080/01431161.2012.674230.

Omernik, M. J. (2004). Perspectives on the nature and definition of ecological regions. *Environmental Management, 34*(1), S27–S38. doi:10.1007/s00267-003-5197-2.

Riemann, R., Wilson, B. T., Lister, A., & Parks, S. (2010). An effective assessment protocol for continuous geospatial datasets of forest characteristics using USFS Forest Inventory and Analysis (FIA) data. *Remote Sensing of Environment, 114*, 2337–2352.

Reyer, C. P., Rammig, A., Brouwers, N., & Langerwisch, F. (2015). Forest resilience, tipping points and global change processes. *Journal of Ecology, 103*(1), 1–4.

Running, S. W., Justice, C. O., Salomonson, V., Hall, D., Barker, J., Kaufmann, Y. J., et al. (1994). Terrestrial remote sensing science and algorithms planned for EOS/MODIS. *International Journal of Remote Sensing, 15*(17), 3587–3620. doi:10.1080/01431169408954346.

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *Bioscience, 54*(6), 547–560. doi:10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2.

Schaphoff, S., Reyer, C. P., Schepaschenko, D., Gerten, D., & Shvidenko, A. (2016). Tamm review: observed and projected climate change impacts on Russia's forests and its carbon balance. *Forest Ecology and Management, 361*, 432–444.

Sha, Z., Bai, Y., Xie, Y., Yu, M., & Zhang, L. (2008). Using a hybrid fuzzy classifier (HFC) to map typical grassland vegetation in Xilin River Basin, Inner Mongolia, China. *International Journal of Remote Sensing, 29*(8), 2317–2337.

Stehman, S. V., Olofsson, P., Woodcock, C. E., Herold, M., & Friedl, M. A. (2012). A global land-cover validation data set, II: augmenting a stratified sampling design to estimate accuracy by region and land-cover class. *International Journal of Remote Sensing, 33*(22), 6975–6993. doi:10.1080/01431161.2012.695092.

Townshend, J. R. G., & Justice, C. O. (2002). Towards operational monitoring of terrestrial systems by moderate-resolution remote sensing. *Remote Sensing of Environment, 83*(1–2), 351–359. doi:10.1016/S0034-4257(02)00082-2.

White, D., Kimerling, J. A., & Overton, S. W. (1992). Cartographic and geometric components of a global sampling design for environmental monitoring. *Cartography and geographic information systems, 19*(1), 5–22.

Woodall, C. W., Oswalt, C. M., Westfall, J. A., Perry, C. H., Nelson, M. D., & Finley, A. O. (2009). An indicator of tree migration in forests of the eastern United States. *Forest Ecology and Management, 257*(5), 1434–1444. doi:10.1016/j.foreco.2008.12.013.

Wright, C. K., & Wimberly, M. C. (2013). Recent land use change in the Western Corn Belt threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences of the United States of America, 110*(10), 4134–4139. doi:10.1073/pnas.1215404110.

Zhang, J., & Foody, G. (1998). A fuzzy classification of sub-urban land cover from remotely sensed imagery. *International Journal of Remote Sensing, 19*(14), 2721–2738.