

Journal of
Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

**Mapping hardwood forests through a
two-stage unsupervised classification
by integrating Landsat Thematic
Mapper and forest inventory data**

Gang Shao
Benjamin P. Pauli
G. Scott Haulton
Patrick A. Zollner
Guofan Shao

Mapping hardwood forests through a two-stage unsupervised classification by integrating Landsat Thematic Mapper and forest inventory data

Gang Shao,^{a,*} Benjamin P. Pauli,^a G. Scott Haulton,^b Patrick A. Zollner,^a and Guofan Shao^a

^aPurdue University, Department of Forestry and Natural Resources, 195 Marsteller Street, West Lafayette, Indiana 47906, United States

^bIndiana Department of Natural Resources, 402 West Washington Street, Indianapolis, Indiana 46204, United States

Abstract. Sound forest management requires accurate forest maps at an appropriate scale. Forest cover data developed at a national scale may be too coarse for forest management at a local level. We demonstrated a two-stage unsupervised classification, integrating Continuous Forest Inventory (CFI) data and Landsat imageries, to classify forest types for Indiana State Forests (ISF) and 8-km surrounding areas. In the first stage, an automatic unsupervised classification assisted by CFI data was applied in ISF. In the second stage, the resultant forest cover information from the first stage was used to expand the classification area into the 8-km surrounding areas. Splitting the classification procedure into two stages made it possible to expand the classification area beyond the coverage of the CFI data. This data-aided unsupervised classification approach increased the repeatability of forest mapping. The resultant map contains five forest types: conifer, conifer-hardwood, maple, mixed hardwood, and oak-hickory forests. The overall accuracy was 81.9%, and the total disagreement was 0.176. The accuracies of conifer, conifer-hardwood, maple, mixed hardwood, and oak-hickory forests were 81.6, 63.4, 75.0, 33.3, and 90%, respectively. This forest mapping technique is suitable for automated mapping of forest areas where extensive plot data are available. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.8.083546](https://doi.org/10.1117/1.JRS.8.083546)]

Keywords: remote sensing; tree species; forest inventory; forest cover; central hardwood forest.

Paper 14328 received Jun. 6, 2014; revised manuscript received Sep. 9, 2014; accepted for publication Sep. 17, 2014; published online Oct. 15, 2014.

1 Introduction

Remote sensing is an effective technology to extract and map spatially explicit information of land use and cover at different scales.^{1,2} Forest cover maps are one of the products made using remotely sensed data and are essential for forest management, habitat monitoring, and biodiversity analysis.³⁻⁶ General-purpose land cover maps developed at national and global scales usually describe forest information at a coarse level. For example, the National Land Cover Database (NLCD)⁷ for the United States contains three forest types: deciduous, evergreen, and mixed forests, a level-II classification defined by the United State Geological Survey (USGS).⁸ For the purposes of intensive forest management, habitat characterization, and forest health monitoring, it is essential to obtain more detailed forest information than the USGS (or Anderson) level-II classification can provide.⁹⁻¹¹

The major challenge in land use or forest classification is to increase classification detail with satisfactory accuracy.¹²⁻¹⁴ This is because forest classification can theoretically be made at any level but classification accuracy decreases with increasing classification detail. Fine-level forest cover maps with low classification accuracy are no better than coarse-level forest cover maps

*Address all correspondence to: Gang Shao, gshao@purdue.edu

with high classification accuracy though different forest management activities may require different forest cover maps. A three-forest-class (or Anderson's level-II) land cover map works well for national-scale forest assessment, but is insufficient for complex forest management at the state level, such as in Indiana where 95% of forests are deciduous.¹⁵ In this region, distinct deciduous hardwood forest types exist, requiring differing silvicultural regimes for management and maintenance.^{10,16} However, detailed classification of hardwood forests is difficult due to the similarities in spectral reflectance, canopy structure, and spatial mixture of hardwood tree species.¹⁷

There is a clear need for a quality forest cover map in Indiana to assess the habitat availability for bat species that are at risk from white-nose syndrome.¹⁸ For example, the Indiana bat (*Myotis sodalis*) is listed as an endangered species by the U.S. Fish and Wildlife Service as well as the International Union for Conservation of Nature.¹⁹ Populations of Indiana bats have declined >50% from 1960 to 2001²⁰ and are under considerable threat of increased declines due to white-nose syndrome.²¹ During the summer, Indiana bats mainly use hardwood and hardwood-pine forests²² and a quality forest cover map can assist in modeling current and future bat habitat.²³

There are several existing forest cover maps that contain more than three forest types at national or state scales. The Forest Cover Types map produced by the National Atlas of the United States²⁴ provides an alternative resource for obtaining a forest cover map in Indiana. This dataset was created based on advanced very high resolution radiometer and Landsat Thematic Mapper (TM) imagery acquired in 1991 and includes 25 types of forests with a 1-km spatial resolution. The classifications of hardwood forests, however, do not separate maple (*Acer* sp.) from mixed hardwood tree species groups despite the fact that maple is one of the most dominant tree genera in Indiana. Furthermore, there is no accuracy assessment information associated with the metadata. The Indiana Gap Analysis Project has also produced a land cover map.²⁵ The development of this land cover map focused upon habitat attributes but did not consider subclasses of hardwood forests. The overall accuracy of this map was just 70.98%. However, for the assessment of land use and land cover mapping, the USGS proposed a recommendation of minimum accuracy of 85%.⁸ For application in landscape quantification, a classification accuracy of >85% is necessary.¹³ Although there is no specific accuracy requirement for the purpose of forest managements and habitat delineation, our objective was to develop a forest cover map with 85% or better classification accuracy. Existing land cover maps were not satisfactory in terms of classification details and accuracy; therefore, it was necessary to develop a more accurate and more detailed forest cover map to meet the needs of various spatial analyses, such as wildlife habitat simulations.

The algorithms of image data classification to derive a forest cover map generally include supervised and unsupervised approaches.²⁶ Supervised classifications require training a dataset to predefine the statistical parameters (such as maximum likelihood) or nonparametric statistical learning functions (such as neural network and support vector machine).^{27,28} Unsupervised classifications (such as ISODATA and K-means clustering) generate spectral clusters based on the statistical information of the remote sensing imagery.^{29,30} Supervised and unsupervised classifications each have their own strengths: the former involves more human input than the latter, whereas the latter is more repeatable than the former. Strictly speaking, both approaches require human experience and neither is absolutely repeatable in practice. It is ideal for a classification algorithm to have minimal human input while maintaining a high level of accuracy and repeatability.³⁰⁻³³ Lang et al.³⁰ demonstrated a data-aided unsupervised classification (DUC) method that interfaced with sample data for labeling spectral classes into information classes. This automated classification approach is superior to the traditional supervised and unsupervised classification methods in terms of ease of use, classification accuracy, and repeatability. However, the use of DUC becomes impractical if there are insufficient ground sample data. Our access to thousands of geo-referenced plots of forest inventory from Indiana State Forests (ISF) provides us with an excellent opportunity to classify forest types with the DUC method. This study represents the first application of DUC to reveal insight about the usefulness of this automated approach in practice.

2 Study Area and Data Processing

This study focused on ISF and the 8-km surrounding areas (37°57'N to 40°54'N, 85°28' W to 87°38'W). Eight-kilometer surrounding areas were implemented because they encompass the

majority of Indiana bat movement from roost sites to foraging areas.^{34,35} A total of 13 state forests with a total area of 9854 km² were included in the study area (Fig. 1). Located in the central hardwood forest region, the forests in the study area are dominated by hardwood species, such as oak (*Quercus sp.*), hickory (*Carya sp.*), maple, and tulip poplar (*Liriodendron tulipifera*).

We downloaded cloud-free Landsat 5 TM data of paths 21 and 22, and rows 32, 33, and 34, acquired in April 2006, September 2008, and October 2010 from the USGS Earth Explorer data website (Table 1).³⁶ There were limited stand-replacing commercial harvests within the study

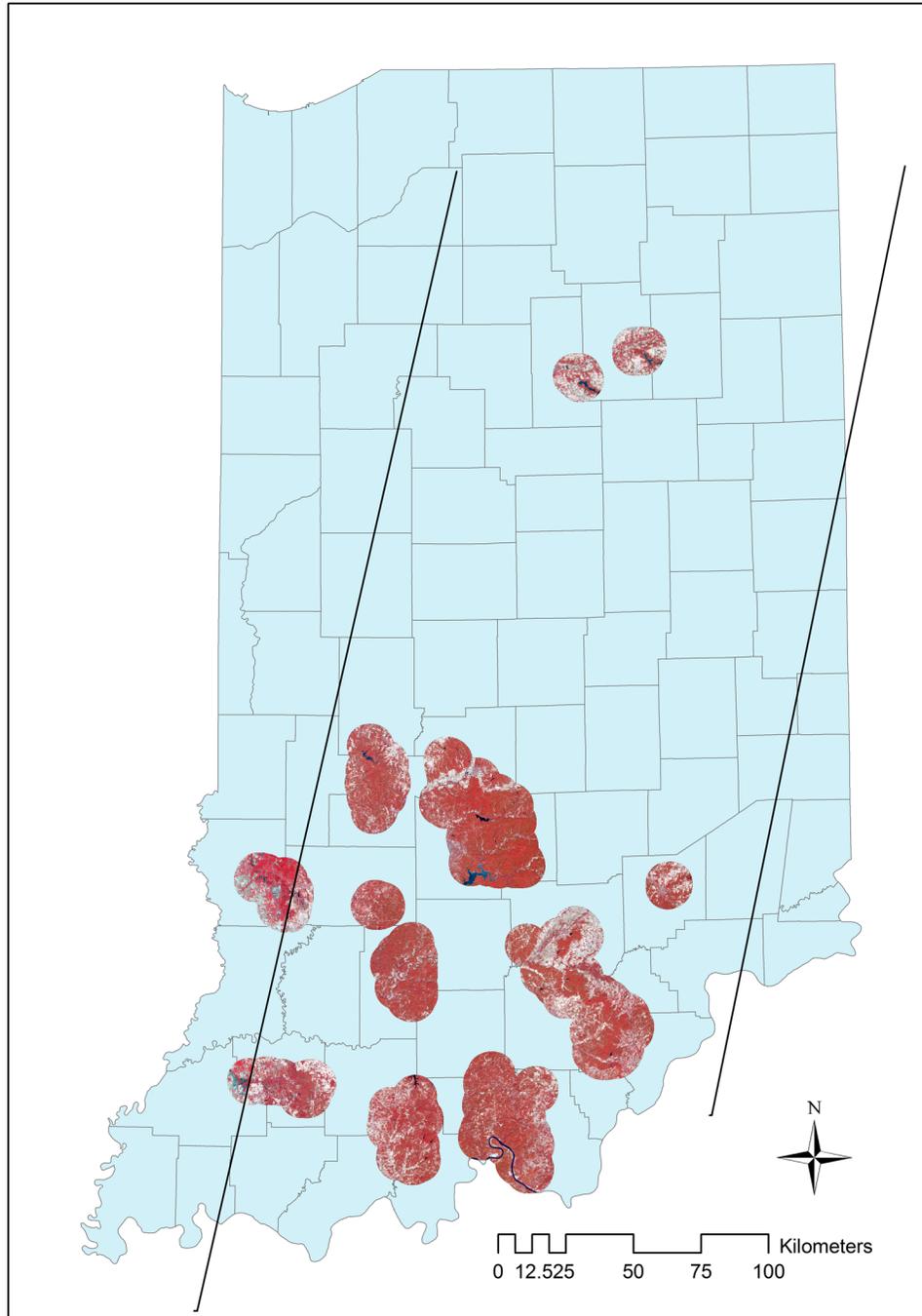


Fig. 1 The extents of Indiana State Forests (ISF) and the 8-km surrounding areas across Indiana. The area between the lines is covered by the Landsat TM path 21 constituting ~97% of the study area. The Landsat TM data were placed as the background with the RGB combination of bands 4, 3, and 2.

Table 1 Landsat 5 Thematic Mapper data used in this study.

Year	Date	Path	Row	Cloud coverage
2006	April 28	21	32, 33, 24	0%
2008	September 24	21	32, 33, 24	0%
2010	October 16	21	32, 33, 24	0%
2010	October 7	22	33	0%

area between 2006 and 2010. A comparison between NLCD 2006 and NLCD 2011 within the study area indicated that only 0.114% of forestland had changed on ISF and the 8-km surrounding areas. Therefore, the overall forest canopy structure remained generally intact during this period. These satellite image data capture spectral characteristics of tree species in spring and fall seasons. The use of different seasons of remotely sensed data has proven useful for improving classification accuracy for mapping tree species in temperate deciduous woodlands.³⁷ In the spring, the deciduous trees have varying levels of greenness due to differences in leaf-growing stages. In the fall, discrimination among deciduous species is possible due to differing colors and amounts of leaves. Therefore, the combination of the two-season datasets increases the ability to distinguish hardwood forest types in Indiana. Mosaics were made from images acquired on the same date and were all clipped with the boundary of our study area. Spring and fall TM image datasets were combined into a single dataset, resulting in two datasets covering 97 and 3% of the area by path 21 and path 22, respectively (Fig. 1). The data-aided classification method (discussed below) was applied to the classification of the path 21 image mosaic. The path 22 image mosaic was classified with a traditional unsupervised classification method due to insufficient plot data available for this small area.²⁶

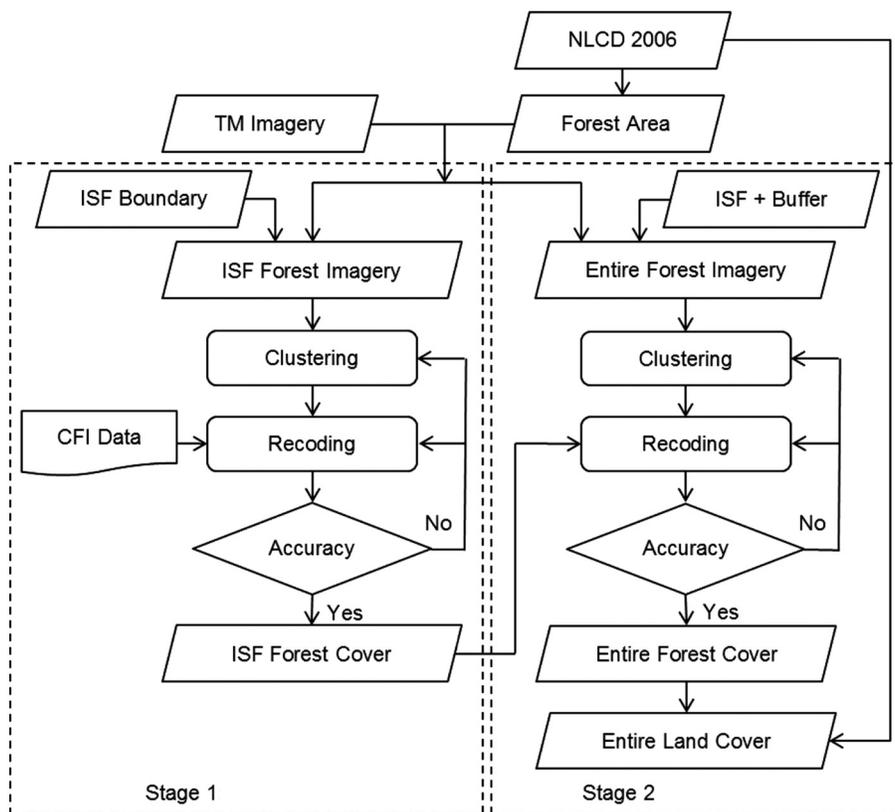


Fig. 2 A flow chart of applying the two-stage, data-aided unsupervised classification method in forest cover mapping for ISFs and the 8-km surrounding areas.

NLCD 2006 was used as a base map, which helped separate forest from nonforested areas (Fig. 2). This ensured the consistency between the newly developed land cover map and NLCD 2006 data for the nonforest categories. We clipped the image mosaics with forest polygons and derived image datasets that contained only forest pixels (Fig. 2).

Continuous Forest Inventory (CFI) plot data were collected on ISF by Indiana Department of Natural Resources (IDNR) between 2006 and 2010 (Fig. 3).³⁸ The sampling intensity was one plot for approximately every 40 acres. Each plot is in a circular shape with a radius of 7.3 m. Stand type was recorded and trees with a diameter at breast height of 5 in. and larger were measured on each plot. In this study, they were used as reference data ($n = 2158$) for image classification (Table 2). We grouped all the CFI plots into five forest types: conifer forest, conifer-hardwood forest, maple forest, mixed hardwood forest, and oak-hickory forest based upon

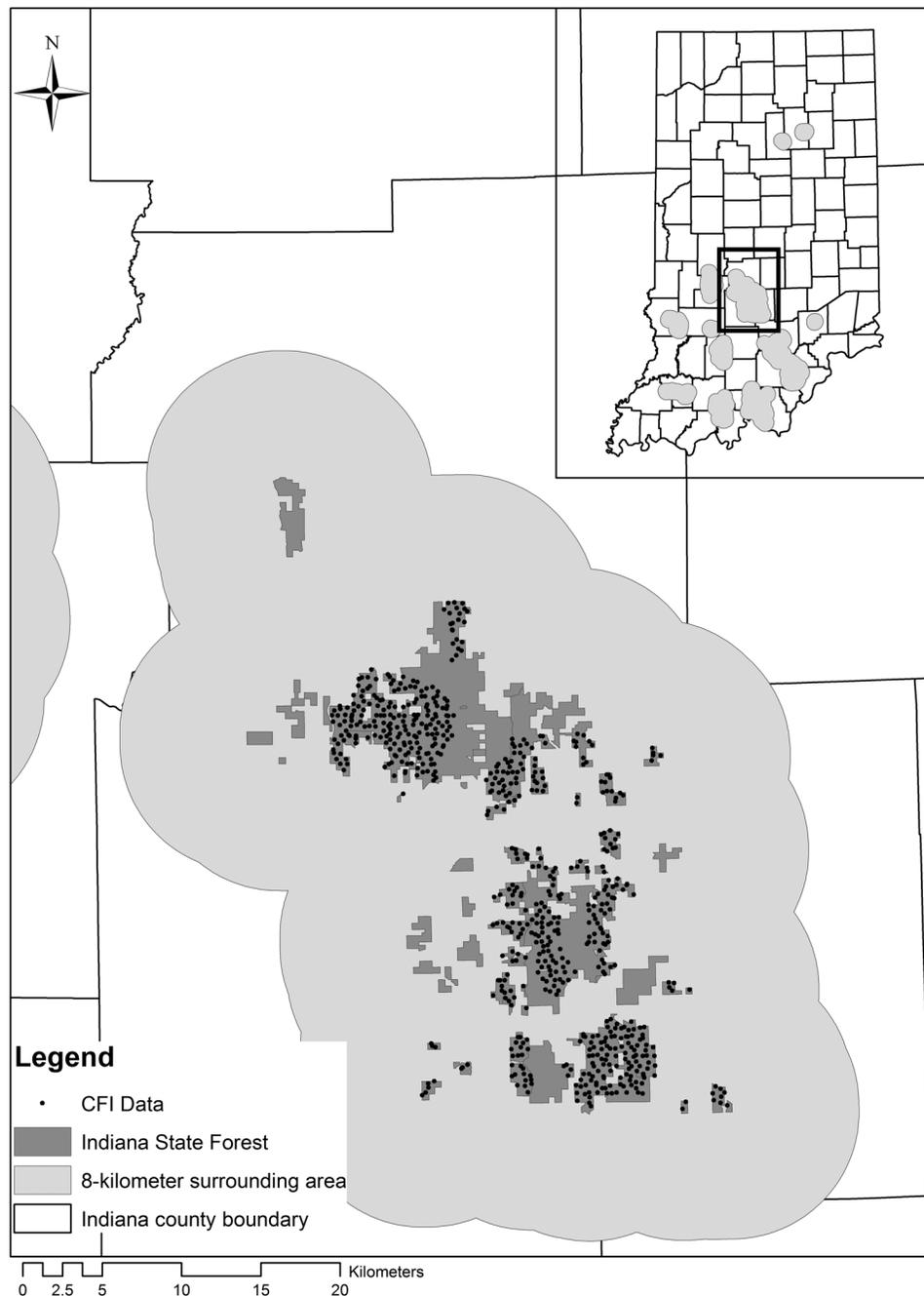


Fig. 3 The spatial distribution of the Continuous Forest Inventory plots in ISF.

the surveyors' forest type classification. The coniferous forest type included all the conifer-dominated plots. If the plot contained both hardwood and coniferous tree species, it was grouped into conifer-hardwood forest type. Maple forest plots were those dominated by maple trees. Plots would belong to oak-hickory forest type if they were dominated by oak and co-dominated by hickory. Mixed hardwood forest plots were those dominated with hardwood tree species other than maple, oak, or hickory. We collected data from an additional 321 plots for accuracy assessment^{39,40} based on the CFI technique above, and 248 of these additional plots were located within ISF area. The 8-km surrounding areas were much greater than ISF in area, but contained only a small portion of reference plots due to difficult accessibility to private forestland.

3 Classification Methods

Band selection is required to improve efficiency and reduce redundancy. Different bands of Landsat 5 TM data have different features for forest mapping.^{41,42} Band 1 has the ability to distinguish deciduous from coniferous vegetation. Band 2 is useful for assessing plant vigor. Band 3 is sensitive to vegetation slopes. Band 4 is related to forest biomass content. Bands 5, 6, and 7 reflect the moisture content of soil and vegetation.⁴³ We performed band-by-band visual examinations to eliminate bands with obvious noise due to high scattered energy and excluded band 6 because of its large pixel size. Only one band would be kept if some bands provide similar information for forest mapping to reduce redundancy.

The TM data analysis and classification were performed in Erdas Imagine 2010 (Ref. 44) and MATLAB® 2010a. The classification procedure included two stages (Fig. 2).

1. We classified ISF by applying the original DUC algorithm.³⁰ Unsupervised clustering was used to achieve preliminary spectral clusters, followed by CFI data assisted cluster labeling. Among 2158 CFI sample plots we used, 1912 were recorded as oak-hickory forest type (Table 2), which dominated the forestland in ISF. Maple, beech (*Fagus sp.*), basswood (*Tilia sp.*), and some other hardwood species are usually intermediate or over-topped in the forest stand.⁴⁵ We initially separated oak-hickory forest from the other forest types. CFI data were used to guide the recoding of oak-hickory forests based on the majority rule, meaning that if over half of the plots belonged to oak and hickory, this cluster would be labeled as oak-hickory forest (Table 2). Due to the extremely uneven distribution of the CFI data among different forest types, a large number of spectral clusters were required to distinguish oak-hickory forest and those forest types with rare CFI data. This procedure resulted in a forest cover map with two forest types: oak-hickory and other forests. Then, other forest types were split using the non-oak-hickory CFI plot data. For the classification of non-oak-hickory forest types, we used the relative-dominant rule: a spectral class was assigned to a forest type that had the most plots within the spectral cluster. If they had equal numbers of plots between any two hardwood plots, the spectral class was assigned to mixed hardwood forest if a cluster contained no coniferous plot. The classification at this stage resulted in a forest cover map within ISF.
2. Both ISF and the 8-km surrounding areas were considered at this stage, but a similar clustering procedure was applied. A large number of spectral clusters were required to identify the forest types with rare CFI data. However, the CFI data were insufficient to label large numbers of spectral clusters in the 8-km surrounding areas using original DUC algorithm because most of the CFI plots were located within the ISF. The clusters that did not overlap with CFI data could not be recoded or labeled with CFI data (Fig. 3). Instead of using CFI plot data for recoding, the resultant forest cover map for ISF from

Table 2 The distribution of Continuous Forest Inventory plots among forest types in Indiana State Forest.

Forest type	Conifer	Conifer-hardwood	Maple	Mixed hardwood	Oak-hickory
Number of points	68	117	16	45	1912

the first classification stage was used in this procedure. Each pixel of this forest cover map was used to label the spectral clusters into forest cover classes by the majority rule of the data-assisted unsupervised classification as discussed in stage 1. Such clustering-recoding was repeated as the number of spectral clusters increased until classification accuracy reached the maximum, resulting in a forest cover map for the entire study area.

Quantity, allocation, and total disagreements, which were developed by Pontius and Millones,⁴⁶ were calculated to qualify the classification results statistically in each stage. Quantity disagreement is defined as the amount of difference in the proportions of the categories between the resultant map and the reference data. Allocation disagreement is the difference in spatial allocation of the categories between the resultant map and the reference data. Total disagreement is the sum of quantity and allocation disagreements.⁴⁶ McNemar's test was used to test the significant difference between the resultant matrices in stages 1 and 2.^{47,48} The Z score was calculated with the following equation:

$$Z = \frac{|a - b|}{\sqrt{a + b}}, \quad (1)$$

where a was the number of samples that were misclassified by the first stage classification but were correctly classified by the second stage classification, and b was the number of samples that were correctly classified by the first stage classification but were misclassified by the second classification. We assumed the results from the two stages were statistically significant ($p < 0.05$) if the Z value was > 1.96 .

A complete land cover dataset was then produced by overlaying this forest cover map with the water layer derived with the original TM imagery and the existing NLCD 2006. The newly derived forest cover dataset is referred to as a five-forest-class (FFC) map, containing conifer forest, conifer-hardwood forest, maple forest, mixed hardwood forest, and oak-hickory forest.

4 Results

In the band selection procession, we selected nine bands from the Landsat 5 TM data for forest cover mapping, including bands 2, 3, 4, and 5 from the 2006 TM data, bands 4 and 5 from the 2008 TM data, and bands 4, 5, and 7 from the 2010 TM data. Bands 1 of all Landsat data were excluded because of the high scattered energy. Bands 6 were all eliminated due to the large pixel size and because the information they provided was not critical for forest mapping. One out of three of bands 2, 3, and 7 were kept to reduce the redundancy. All bands 4 and 5 were included because they are sensitive to the seasonal changes of the spectral reflectance of the forest species.

To separate oak-hickory forest from other forest types in stage 1, we started with 100 spectral clusters using unsupervised classification and increased the number of the clusters by a 50-cluster interval.²⁶ By adjusting the number of spectral clusters, the highest classification accuracy was reached when the number of spectral clusters was 200. For the classification of non-oak-hickory forest types, the classification accuracy reached the highest when the number of spectral clusters was 100. In this step, there would be individual spectral cluster/clusters that had no overlap with any CFI plot if there were too many spectral clusters. The FFC map within ISF was created at this stage [Fig. 4(a)]. In the second stage, 200 spectral clusters were classified to create an FFC map within ISF and the 8-km surrounding areas [Fig. 4(b)]. The accuracy was expressed with spatial agreement of forest types within shared ISF area between two forest cover maps from both stages.

The FFC map resulting from the first stage of the classifications reached an overall accuracy of 74.2% (Table 3). The quantity and allocation disagreements were 0.052 and 0.206. The total disagreement was 0.258.⁴⁶ The oak-hickory forest type had the highest accuracy on the average of user's and producer's accuracies, followed by conifer, conifer-hardwood, maple, and mixed hardwood forests. The classification accuracies of conifer-hardwood, maple, and mixed hardwood forests were lower than desirable. The second stage of the classifications increased the overall accuracy to 81.9% (Table 4), better than the 78% overall accuracy of NLCD 2006 level II classification [Fig. 4(c)].⁴⁹ The total disagreement was 0.176 with a quantity disagreement of

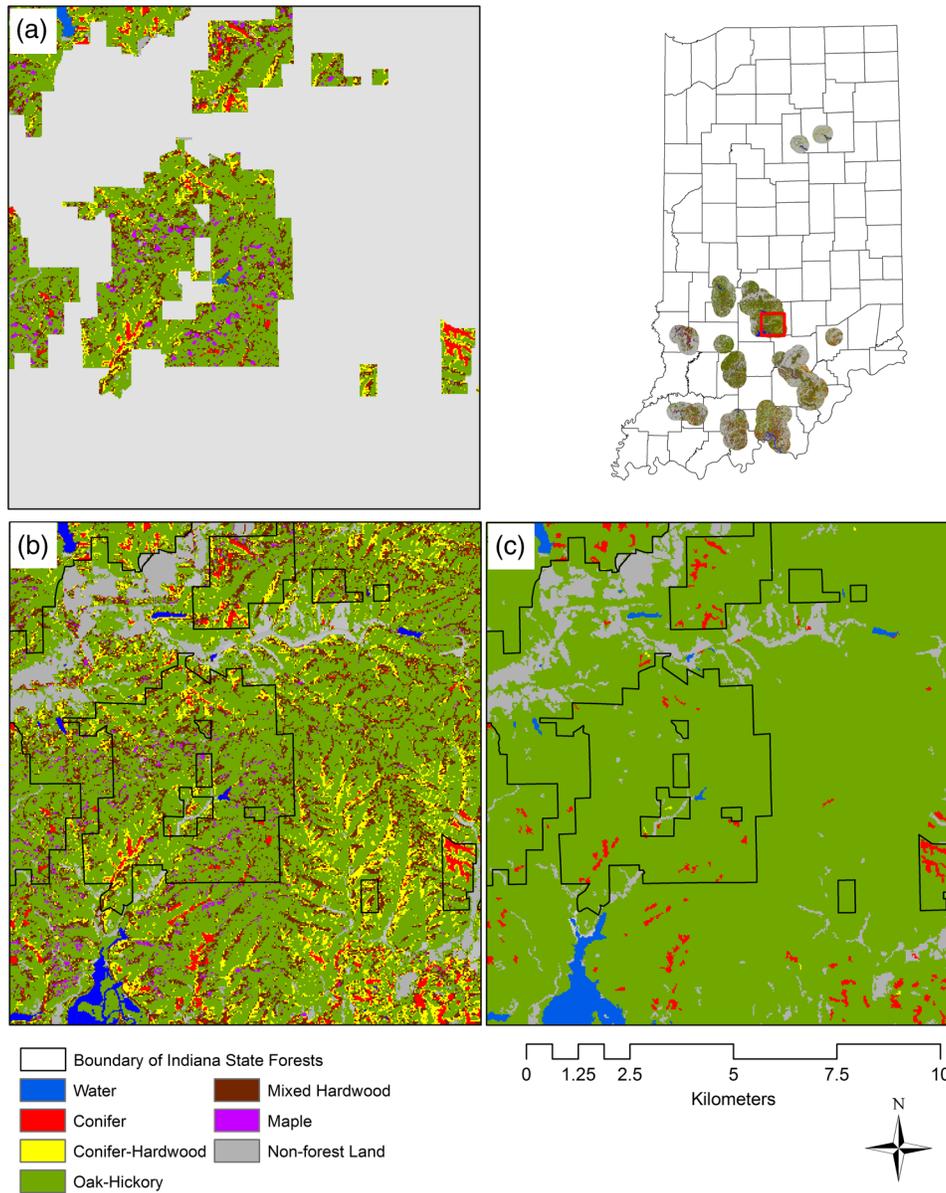


Fig. 4 The ISF forest cover map (a), the forest cover map of the ISF and the 8-km surrounding areas (b), and the National Land Cover Database (NLCD) 2006 (c).

0.028 and allocation disagreement of 0.148. Again, the oak-hickory forest type had the highest accuracy; conifer and maple forest types also had satisfactory accuracies. The mixed hardwood forest type had a rather low producer's accuracy though its user's accuracy was relatively high. McNemar's test was applied to compare the results from the first and second stages of the classification. There were 73 plot samples that were used in the second stage of the classification but were not used in the first stage of the classification because they were located in the 8-km surrounding areas of ISF. As McNemar's test requires the same quantity of samples from these two matrices, these 73 samples were not used in the McNemar's test, among which, 7 samples were misclassified in the second stage of the classification. The Z value of the McNemar's test was 2.98, which was > 1.96 . Therefore, the difference of the results from the two classification stages was statistically significant at a 96% confidence level, which means that the result in the second stage of classification was significantly improved.

The FFC map shows the dominant compositions of oak-hickory and mixed hardwood forests in forest landscape, comprising 81% of the forested landscape in areas within ISF [Fig. 5(a)].

Table 3 An error matrix of classification for forests within Indiana State Forest following the stage one classification.

Classification	Reference					Total	UA (%)
	Conifer	Conifer-hardwood	Maple	Mixed hardwood	Oak-hickory		
Conifer	39	6	0	1	1	47	83.0
Conifer-hardwood	1	19	0	2	7	29	65.5
Maple	1	0	14	0	11	26	53.9
Mixed hardwood	0	2	0	6	2	10	60.0
Oak-hickory	8	14	2	6	106	136	82.8
Total	49	41	16	15	127	248	
PA (%)	79.6	46.3	87.5	40.0	83.5		OA = 74.2

Note: UA, user's accuracy; PA, producer's accuracy; OA, overall accuracy.

Table 4 An error matrix of classification for forests in Indiana State Forest and the 8-km surrounding areas following the stage two classification.

Classification	Reference					Total	UA (%)
	Conifer	Conifer-hardwood	Maple	Mixed hardwood	Oak-hickory		
Conifer	40	4	0	1	2	47	85.1
Conifer-hardwood	1	26	0	5	15	47	55.3
Maple	1	0	12	0	2	15	80.0
Mixed hardwood	0	0	0	5	1	6	83.3
Oak-hickory	7	11	4	4	180	206	87.3
Total	49	41	16	15	200	321	
PA (%)	81.6	63.4	75.0	33.3	90.0		OA = 81.9

Note: UA, user's accuracy; PA, producer's accuracy; OA, overall accuracy.

This result was consistent with forest composition data reported by IDNR (2008),¹⁰ which were obtained from field observations. Maple forest type in the FFC map constitutes 7% of ISF area, equivalent to the sum of 4% for maple and 3% for bottomland hardwoods (IDNR 2008) [Fig. 5(b)]. The combined area of conifer and conifer-hardwood mixed forests in the FFC map was slightly more than the pine forest area reported by IDNR (2008). The proportion of conifer-hardwood forest in the FFC map was reasonable when compared with independent forest inventory and analysis (FIA) data from USDA Forest Service.⁵⁰ The conifer-hardwood forest covered a much larger area in the FFC map than in NLCD 2006 data either inside or outside ISF and is much more detailed than the forest cover classifications in NLCD 2006 [Fig. 5(c)].

5 Discussion

This study is the first demonstration of the DUC algorithm³⁰ for a real-world classification exercise with an integration of remotely sensed and plot-based forest data. Lang et al.³⁰ designed and tested the DUC algorithm to classify several sample sites into four general land covers, including agriculture, forest, urban, and water. In this study, we applied this algorithm to a much larger site,

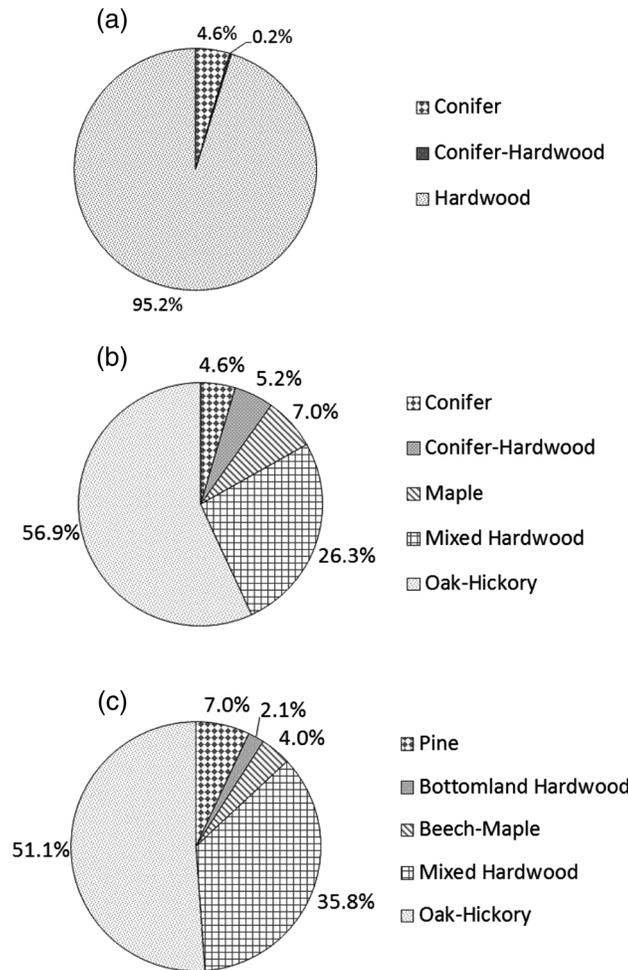


Fig. 5 A comparison of forest composition in area among NLCD 2006 (a), five-forest-class map (b), and report by IDNR (2008) (c) for ISF.

the ISF with the 8-km surrounding areas. We focused on and split the forests into more detailed forest types, which was more difficult than the general land cover classification due to the similarities. The classification accuracy of such an unsupervised classification approach is determined mainly by imagery quality and sample data. The two-season Landsat TM datasets we used in this study provide richer spectral information than single-season imagery for the classifications of hardwood forests. The CFI plots used in this study were abundant though their distribution among forest types was uneven. All these factors contributed to the development of the FFC map that is reasonably consistent with field survey data. This shows that the integration of image data and forest field data has made the resultant forest cover map more realistic and objective than the use of image data alone.

The overall accuracy of the forest cover map derived at the second stage was unexpectedly greater than that at the first stage. The reason for this phenomenon may be that spectral classes derived from a larger forest area (ISF and the 8-km surrounding areas) have better representations to forest types than those from a smaller forest area (ISF). Because most sample points are located inside ISF, forest cover information within ISF may be more reliable than that outside ISF. The producer's accuracy of maple forest type in the second stage was lower than that in the first stage; however, the user's accuracy had significant increases. This means the FFC map created in the first stage overestimated the maple. The FFC of the second stage showed a lower quantity disagreement. It is also reasonable for the additional misclassified maple plots dropping into oak-hickory forest type because maples are shade tolerance species and are usually suppressed under the oak-hickory forests. The same trend of accuracies happened in mixed hardwood forest, where the producer's accuracy decreased and the user's accuracy

increased. However, there was no confidential evidence to show that mixed hardwood was overestimated in the first stage. FFC in the second stage might have underestimated the mixed hardwood because the total number of mixed hardwood samples in the resultant map was much smaller than that in the reference data. The decrease of the allocation disagreement showed that FFC in the second stage would likely have a better quality.

The classification accuracy is also affected, in theory, by spatial mismatch between pixel coordinates and plot locations. The coordinates of the CFI plots used in this study were measured with handheld global positioning system (GPS) receivers and their errors are normally up to 10 m.⁵¹ The combination of the GPS error and TM image geometric error can be problematic for heterogeneous forest canopies. The coverage of this type of forest inventory data is available for limited forest areas, geographically restricting the applications of the data-assisted unsupervised classifications.

Given the first experiment with the classification method, we suggest that it should be broadly applied with the following considerations:

1. Temporal consistency: The time of image data to be used for classification should be close to the time of forest data collection. If there is a difference in time, areas that have changed between the two times should be excluded from image data classification.
2. Spatial correction: Both image data and plot data should be correctly geo-referenced. Geometric corrections may be needed though most remotely sensed data have been geo-referenced by the data provider. It is ideal that forest plot coordinates measured with GPS are differentially corrected. Because the data-aided unsupervised classification is a nonparametric approach, sample plots for classification can be purposely located in the middle of forest stands on the ground. Alternatively, plots located on the edges of forest stands can be excluded from their use for assisting image data classification.
3. Spatial representation: Sample plots should have an extensive coverage over the study area and, thus, only the first stage is needed for completing image data classification. Plot size needs to be big enough to have a good representation of tree composition from which forest types are derived. If the plot data are used exclusively for image data classification, the simplest attribute measurement is to record the name of the forest type by experienced forest surveyors. Only the canopy tree species is helpful to image data classification.
4. Map validation: The overall accuracy is the first but not the only consideration for the quality of a forest cover map. It is essential that a forest cover map be assessed with plot data that are not used for assisting image data classification.³⁹ The sample plots that are used for accuracy assessment should be randomly located on the ground. It is best to have a broad coverage over the study area with reference sample plots. It is important to balance between the user's and producer's accuracies.^{13,52}

An important feature of the DUC algorithm is its repeatability, meaning that the classification procedure can be systematically modified by simply changing classification parameters and labeling rules if classification results are not satisfactory, thus the loops in the flow chart of the two-stage unsupervised classification (Fig. 2). Our experience indicates that an automated computer program for recoding can make the classification less labor intensive and reduce human errors in labeling processes despite hundreds of spectral classes. Analysts only need to have fundamental remote sensing training to implement this classification technique.

Forest inventory data, such as FIA, have been used in various ways to improve forest assessment together with remotely sensed imagery.^{40,53,54} If researchers can access the exact FIA plot locations, the FIA-DUC approach can be broadly applied at the regional and national scales in the United States. Such a forest mapping procedure will help save time and money due to a reduction in the necessary ground validation.

6 Conclusions

This study demonstrates the first application of the DUC algorithm in dividing hardwood forests into three forest types in an objective fashion. The overall quality of the resultant forest cover map suggests that the DUC approach with forest inventory data is an effective and efficient

method for mapping forest cover in Indiana. However, current ground data do not allow us to classify the hardwood forest into even more detailed levels with satisfactory accuracy. A step-wise classification procedure of species composition overcomes the difficulties caused by the extremely uneven distribution of ground data. The two-stage DUC algorithm successfully extends the mapping area to 8 km away from the plot-based forest data without jeopardizing classification accuracy. This forest mapping technique is suitable for mapping other forest areas where extensive plot data are available.

A forest cover map needs to be noise free if it is to be used for forest management activities in the field. In other words, the minimum mapping unit will likely be greater than pixel size of the remote sensing imagery at a proper scale based on the objective of the map to reduce salt-and-pepper or noise pixels. A forest cover map derived from pixel-based classifications can be filtered to remove noise pixels by using image processing algorithms, such as connected component analysis and morphology fundamentals processing. It is also possible to integrate spectral classes with image segmentation to obtain patch-style spectral classes, based on which labeling procedure is implemented. Such comprehensive experiments need systematic explorations in the future.

References

1. S. E. Franklin and M. A. Wulder, "Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas," *Prog. Phys. Geogr.* **26**(2) 173–205 (2002).
2. S. D. Jawak and A. J. Luis, "Improved land cover mapping using high resolution multiangle 8-band worldview-2 satellite remote sensing data," *J. Appl. Remote Sens.* **7**(1), 073573 (2013).
3. W. B. Cohen et al., "Modelling forest cover attributes as continuous variables in a regional context with Thematic Mapper data," *Int. J. Remote Sens.* **22**(12), 2279–2310 (2001).
4. J. L. Ohmann and M. J. Gregory, "Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA," *Can. J. For. Res.* **32**(4), 725–741 (2002).
5. K. M. Bergen et al., "Remote sensing of vegetation 3-D structure for biodiversity and habitat: review and implications for lidar and radar spaceborne missions," *J. Geophys. Res.: Biogeosci.* **114**(G2), 2005–2012 (2009).
6. E. H. Helmer et al., "Detailed maps of tropical forest types are within reach: forest tree communities for Trinidad and Tobago mapped with multiseason Landsat and multiseason fine-resolution imagery," *For. Ecol. Manage.* **279**(1), 147–166 (2012).
7. J. A. Fry et al., "Completion of the 2006 National Land Cover Database for the Conterminous United States," *Photogramm. Eng. Remote Sens.* **77**(9), 859–864 (2011).
8. J. R. Anderson, E. E. Hardy, and J. T. Roach, "A land use and land cover classification system for use with remote sensor data," Vol. 964, US Government Printing Office (1976).
9. D. R. Breininger et al., "Landcover characterizations and Florida scrub-jay (*Aphelocoma coerulescens*) population dynamics," *Biol. Conserv.* **128**(2), 169–181 (2006).
10. Indiana Department of Natural Resources, Division of Forestry, "Increased emphasis on management and sustainability of oak-hickory communities on the Indiana State forest system, 2008–2027," http://www.in.gov/dnr/forestry/files/fo-StateForests_EA.pdf (2008).
11. R. K. Swihart et al., "The hardwood ecosystem experiment: a framework for studying responses to forest management," <http://www.nrs.fs.fed.us/pubs/42882> (2013).
12. G. M. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.* **80**(1), 185–201 (2002).
13. G. F. Shao and J. G. Wu, "On the accuracy of landscape pattern analysis using remote sensing data," *Landsc. Ecol.* **23**(1), 505–511 (2008).
14. S. Adelabu et al., "Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image," *J. Appl. Remote Sens.* **7**(1), 073480 (2013).
15. J. Gallion, "Indiana's forest resource in 2011," http://www.inwoodlands.org/indianas_forest_resource_in_20/ (2011).
16. J. Gallion and C. Woodall, "The sustainability of Indiana's forest resources," [http://www.in.gov/dnr/forestry/files/fo-SIFR\(lowres\).pdf](http://www.in.gov/dnr/forestry/files/fo-SIFR(lowres).pdf) (2010).

17. R. Pu, "Broadleaf species recognition with in situ hyperspectral data," *Int. J. Remote Sens.* **30**(11), 2759–2779 (2009).
18. W. E. Thogmartin et al., "Population-level impact of white-nose syndrome on the endangered Indiana bat," *J. Mammal.* **93**(4), 1086–1098 (2012).
19. *Myotis sodalis*, The IUCN red list of threatened species, <http://www.iucnredlist.org/details/14136/0>.
20. R. L. Clawson, "Trends in population size and current status," in *The Indiana Bat: Biology and Management of an Endangered Species*, A. Kurta and J. Kennedy, Eds., pp. 2–8, Bat Conservation International, Austin, TX (2002).
21. W. E. Thogmartin et al., "White-nose syndrome is likely to extirpate the endangered Indiana bat over large parts of its range," *Biol. Conserv.* **160**(April), 162–172 (2013).
22. E. V. Callahan, R. D. Drobney, and R. L. Clawson, "Selection of summer roosting sites by Indiana bat (*Myotis sodalis*) in Missouri," *J. Mammal.* **78**(3), 818–825 (1997).
23. B. P. Pauli, "Nocturnal and diurnal habitat of Indiana and Northern long-eared bats, and the simulated effect of timber harvest on habitat suitability," PhD Dissertation, Purdue University, West Lafayette, Indiana (2014).
24. Forest Cover Types Info, <http://www.nationalatlas.gov/mld/foresti.html>.
25. Gap Analysis Bulletin No. 16, March 2009, <http://www.gap.uidaho.edu/bulletins/16/Indiana.pdf>.
26. J. R. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective*, 3rd ed., Prentice Hall, NY (2005).
27. Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitioning clustering techniques," *IEEE Trans. Geosci. Remote Sens.* **47**(8), 2973–2987 (2009).
28. G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: a review," *ISPRS J. Photogramm. Remote Sens.* **66**(3), 247–259 (2011).
29. D. S. Lu and Q. H. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.* **28**(5), 823–870 (2007).
30. R. L. Lang et al., "Optimizing unsupervised classifications of remotely sensed imagery with a data-aided labeling approach," *Comput. Geosci.* **34**(12), 1877–1885 (2008).
31. J. Franklin et al., "Rationale and conceptual framework for classification approaches to assess forest resources and properties," in *Remote Sensing of Forest Environments*, M. A. Wulder and S. E. Franklin, Eds., pp. 279–300 (2003).
32. M. Song, D. L. Civco, and J. D. Hurd, "A competitive pixel-object approach for land cover classification," *Int. J. Remote Sens.* **26**(22), 4981–4997 (2005).
33. B. Schmook et al., "A step-wise land-cover classification of the tropical forests of the Southern Yucatán, Mexico," *Int. J. Remote Sens.* **32**(4), 1139–1164 (2011).
34. S. W. Murray and A. Kurta, "Nocturnal activity of the endangered Indiana bat (*Myotis sodalis*)," *J. Zool.* **262**(2), 197–206 (2004).
35. D. W. Sparks et al., "Foraging habitat of the Indiana bat (*Myotis sodalis*) at an urban–rural interface," *J. Mammal.* **86**(4), 713–718 (2005).
36. EarthExplorer, <http://earthexplorer.usgs.gov/>.
37. R. A. Hill et al., "Mapping tree species in temperate deciduous woodland using time-series multi-spectral data," *Appl. Veg. Sci.* **13**(1), 86–99 (2010).
38. J. Gallion, "Report of Continuous Forest Inventory (CFI)—Summary of year 1–5 (2008–2012)," http://www.in.gov/dnr/forestry/files/fo-CFI_Report_2008-12.pdf.
39. R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.* **37**(1), 35–46 (1991).
40. L. Leefers and N. Subedi, "Forest type classification accuracy assessment for Michigan's State and National Forests," *North. J. Appl. For.* **29**(1), 35–42 (2012).
41. D. S. Lu et al., "Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin," *For. Ecol. Manage.* **198**(1), 149–167 (2004).
42. G. F. Shao et al., "Forest cover types derived from Landsat Thematic Mapper imagery for Changbai Mountain area of China," *Can. J. For. Res.* **26**(2), 206–216 (1996).
43. FAQ about the Landsat Missions, http://landsat.usgs.gov/best_spectral_bands_to_use.php.

44. Erdas Imagine, <http://www.hexagongeospatial.com/products/remote-sensing/erdas-imagine>.
45. P. S. Johnson et al., *The Ecology and Silviculture of Oaks*, 2nd ed., CABI (2009).
46. R. G. Pontius and M. Millones, "Death to kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment," *Int. J. Remote Sens.* **32**(15), 4407–4429 (2011).
47. Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika* **12**(2), 153–157 (1947).
48. A. H. Bowker, "A test for symmetry in contingency tables," *J. Am. Stat. Assoc.* **43**(244), 572–574 (1948).
49. J. D. Wickham et al., "Accuracy assessment of NLCD 2006 land cover and impervious surface," *Remote Sens. Environ.* **130**(15), 294–304 (2013).
50. What types of tree grow in Indiana, <http://na.fs.fed.us/spfo/pubs/misc/in98forests/webversion/whatypes.htm>.
51. M. G. Wing, A. Eklund, and L. D. Kellogg, "Consumer-grade global positioning system (GPS) accuracy and reliability," *J. For.* **103**(4), 169–173 (2005).
52. G. F. Shao et al., "An explicit index for assessing the accuracy of cover class areas," *Photogramm. Eng. Remote Sens.* **69**(8), 907–913 (2003).
53. M. D. Nelson et al., "Combining satellite imagery with forest inventory data to assess damage severity following a major blow down event in northern Minnesota, USA," *Int. J. Remote Sens.* **30**(19), 5089–5108 (2009).
54. Y. Zhang et al., "Integration of satellite imagery and forest inventory in mapping dominant and associated species at a regional scale," *Environ. Manage.* **44**(2), 312–323 (2009).

Gang Shao received his BS in biomedical information engineering from Northeastern University in China and his MS in forestry and natural resources from Purdue University, USA. He is currently a PhD student in forestry and natural resources at Purdue University, USA.

Benjamin P. Pauli received his BA in biology from Lawrence University, USA, and received his MS and PhD in quantitative ecology from Purdue University, USA. He is currently a postdoctoral researcher at Boise State University, USA.

G. Scott Haulton received a BS in English from the State University of New York College at Geneseo, USA, a BS in environmental and forest biology from the State University of New York College of Environmental Science and Forestry, USA, and an MS in wildlife science from Virginia Tech, USA. He is current a forestry wildlife specialist with the Indiana Department of Natural Resources, Division of Forestry.

Patrick A. Zollner received his BS in natural resources from University of Michigan, USA, his MS in wildlife ecology from Mississippi State University, USA, and his PhD in ecology from Indiana State University, USA. He is currently an associate professor of quantitative ecology in the Department of Forestry and Natural Resources, Purdue University, USA.

Guofan Shao received the PhD degree in ecology from the Chinese Academy of Sciences, Shenyang, China, and received postdoctoral education from the Department of Environmental Sciences, University of Virginia, USA. He is currently a professor with the Department of Forestry and Natural Resources, Purdue University, USA. His main research interests focus on remote-sensing data analysis and error propagation, landscape quantification, urbanization processes, and geospatial modeling.