USGS
science for a changing world

Prepared in collaboration with the National Park Service Inventory and Monitoring Program's Gulf Coast Network

Unsupervised Classification of Lidar-based Vegetation Structure Metrics at Jean Lafitte National Historical Park and Preserve

Open-File Report 2012–1096

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia 2012

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth,
its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Suggested citation:
Kranenburg, C., Palaseanu-Lovejoy, M., Nayegandhi, A., Brock, J.C. and Woodman, R., 2012, Unsupervised Classification of Lidar-based Vegetation Structure Metrics at Jean Lafitte National Historical Park and Preserve: U.S. Geological Survey Open-File Report 2012-1096, 19 p.

Acknowledgments

The U.S. Geological Survey Coastal and Marine Geology Program and National Park Service Inventory and Monitoring Program funded this investigation. The authors gratefully acknowledge C. Wayne Wright for Experimental Advanced Airborne Research Lidar (EAARL) instrumentation development and operating the system during this mission.

Contents

Figures

Tables

Conversion Factors

SI to Inch/Pound

Multiply	By	To obtain
Length		
centimeter (cm)	0.3937	inch (in.)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
Area		
hectare (ha)	2.471	acre
square meter (m^2)	10.76	square foot (ft^2)
square kilometer (km^2)	0.3861	square mile (mi^2)

Abbreviations and Acronyms

AGBM	Aboveground Biomass
ALPS	Airborne Lidar Processing System
ASCII	American Standard Code for Information Interchange
BE	Bare Earth
CH	Canopy Height
CIR	Color Infrared
CLARA	Clustering Large Applications
CR	Canopy Reflection
CRR	Canopy Reflection Ratio
DEM	Digital Elevation Model
EAARL	Experimental Advanced Airborne Research Lidar
GeoTIFF	Georeferenced Tagged Image File Format
GR	Ground Reflection
HOME	Height of Median Energy
JELA	Jean Lafitte National Historical Park and Preserve
lidar	Light Detection and Ranging
MCD	Minimum Covariance Determinant
MSL	Mean Sea Level
NASA	National Aeronautics and Space Administration
NPS	National Park Service
NWRC	National Wetlands Research Center
PCA	Principal Component Analysis
RCF	Random Consensus Filter
TIN	Triangulated Irregular Network
UTM	Universal Transverse Mercator

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).
Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

Unsupervised Classification of Lidar-based Vegetation Structure Metrics at Jean Lafitte National Historical Park and Preserve

Prepared in collaboration with the National Park Service Inventory and Monitoring Program's Gulf Coast Network

by Christine Kranenburg,[1] Monica Palaseanu-Lovejoy,[1] Amar Nayegandhi,[1] John Brock,[2] and Robert Woodman[3]

[1]Jacobs Technology, contracted to the U.S. Geological Survey, [2]U.S. Geological Survey, [3]National Park Service

Introduction

The objective of this study is to classify the major types of vegetation assemblages in the Barataria Preserve at Jean Lafitte National Historical Park and Preserve (JELA), Louisiana, using metrics derived from the Experimental Advanced Airborne Research Lidar (EAARL) system. The EAARL is a raster-scanning, waveform-resolving, green-wavelength (532 nm) lidar system designed to map nearshore bathymetry, topography, and vegetation structure simultaneously. The EAARL sensor was developed (circa 2000) by the National Aeronautics and Space Administration (NASA) at its Wallops Flight Facility, Virginia. The EAARL instrument records the time history of the return waveform within a small footprint (20-cm diameter on the ground) for each laser pulse, enabling characterization of vegetation canopy structure and bare earth (BE) topography beneath a variety of vegetation types. The EAARL system also acquires concurrent, high-resolution geolocated color infrared (CIR) imagery at a 1-second interval. A collection of individual waveforms is combined to create a synthesized large-footprint waveform that is used to define three canopy metrics: canopy height (CH), canopy reflection ratio (CRR), and height of median energy (HOME). For this study, a 5-m footprint size was used, but in general, the appropriate size of the synthesized footprint is derived based on a combination of the lidar sampling density and the nature of the terrain (Nayegandhi and others, 2006). The lidar-based vegetation canopy metrics, along with BE elevation data, were then used in an unsupervised classification procedure to determine the boundaries or patches of vegetation structural communities within the Barataria Preserve.

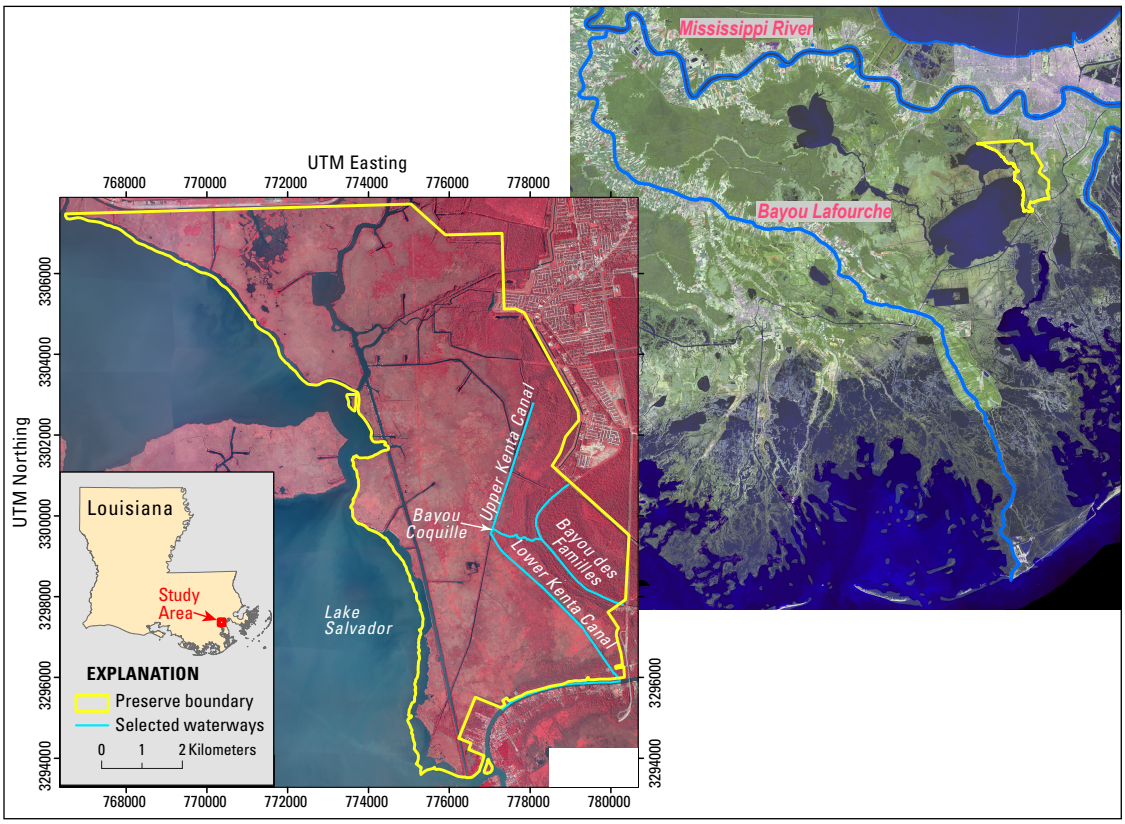


Figure 1. Study area location map with color-infrared (CIR) aerial photograph showing Barataria Preserve boundary and selected waterways.

Study Area

Jean Lafitte National Historical Park and Preserve consists of six separate units across southern Louisiana: The Acadian Cultural Center, the Prairie Acadian Cultural Center, the Wetlands Acadian Cultural Center, the Chalmette Battlefield and National Cemetery, the French Quarter Visitor Center and the Barataria Preserve. The Barataria Preserve unit is located south of New Orleans, Louisiana, and is the Park's most ecologically diverse

site, with over 20,000 acres of swamp, marsh, trails, and waterways (fig. 1). The study area is approximately two-thirds marsh with the remaining area composed of a swamp forest (National Park Service, 2009a).

The National Park Service (NPS) Environmental Assessment report (2009b) describes the geology of the Preserve as being driven by its location relative to the historical path of the Mississippi River. The Barataria Basin, in which the Preserve is located, is bounded on the north and east by the current main stem of the Mississippi River and on the west by Bayou Lafourche. Bayou Lafourche is a distributary located 35 miles west of the current main channel. The report states that the Bayou des Familles, an abandoned distributary that curves through the southeastern section of the Preserve, is flanked on either side by natural levees approximately 1.5 m (5 ft) above mean sea level (MSL).

In their 2004–2007 vascular plant inventory of JELA, Urbatsch, Ferguson, and Gunn-Zumo (2007) documented bottomland hardwood forest, baldcypress-tupelo swamp, shrub-scrub swamp, fresh marsh, intermediate marsh, and flotant marsh as occurring within the preserve. Flotant marshes are defined as floating fresh marshes "composed of vegetation rooted in an organic mat that detaches from the underlying mineral substrate and shifts vertically as water levels below rise and drop" (Battaglia, Denslow, and Hargis, 2007, p. 1).

Approximately two-thirds of the Barataria Preserve consists of a large flat plain of fresh and flotant marsh extending from the western edge of the preserve, bordering Lake Salvador, to the banks of the Kenta Canal (fig. 1) (Battaglia, Denslow, and Hargis, 2007). This fresh marsh is dominated by structurally homogenous vegetation, primarily bulltongue arrowhead (*Sagittaria lancifolia*), but is interrupted by isolated patches of shrub-scrub swamp (Urbatsch, Ferguson, and Gunn-Zumo, 2007). Homer and others (2004, p. 8) define shrub-scrub swamps as "areas dominated by shrubs; less than 5 meters tall with shrub canopy typically greater than 20 percent of total vegetation." They are often found in scattered patches throughout the interior marshes. The fresh marsh plain is also crisscrossed by a large network of dredged canals, each giving rise to spoil banks and their associated vegetation. Spoil banks are artificially created ridges consisting of material dredged from canals that is deposited on either side of the adjacent waterway. The banks are typically 30 m wide and, in contrast with the surrounding marsh, have elevated and drained soil that supports woody vegetation, including many bottomland hardwood forest species (Monte, 1978). Spoil banks are also frequently colonized by invasive exotic species, in particular, Chinese tallow (*Triadica sebifera*) (Urbatsch, Ferguson, and Gunn-Zumo, 2007).

East of the Kenta Canal the vegetation changes from the fresh marsh to that of a baldcypress-tupelo swamp. "[Baldcypress-tupelo swamp] includes forested, alluvial swamps growing on sporadically exposed soils that are generally saturated or inundated throughout most of the growing season, except for periods of extreme drought" (Urbatsch, Ferguson, and Gunn-Zumo, 2007, p. 21). As the name suggests, these swamps are codominated by baldcypress (*Taxodium distichum*) and water tupelo (*Nyssa aquatica*) trees. Both of these species consist of large trees with heights usually in excess of 100 ft (30 m). Structurally, both species are marked by a strongly buttressed base surrounded by one or more pneumatophores (knees) and a long, tapering bole (trunk), but the crown shapes are distinctly different. Baldcypress is characterized by an open and narrowly pyramidal crown, whereas the water tupelo has a flattened, oblong crown containing numerous branchlets. Baldcypress and water tupelo characteristics were obtained from the Trees of Florida database (Institute of Food and Agricultural Sciences, 2011).

In the northern two-thirds of the park (but east of the Kenta Canal), the transition is fairly straightforward. From west to east, marsh grasses change to marsh grasses interspersed with sparse, relatively short trees, and then to taller, more densely populated forest (Woodman, R., personal communication, 2008). In the southeastern section of the Preserve, where it intersects the Bayou des Familles, the transition is more complex. Superimposed on the transition is a ridge of bottomland hardwood forest on the natural levees alongside the bayou. The transition to swamp forest then resumes as one progresses down the backslope of the levee (Denslow and Battaglia, 2002).

Bottomland hardwood forests occur in areas subject to intermittent to frequent flooding (Mitsch and Gosselink, 2007), but that are usually dry. As such, these forests generally appear on relatively higher ground than the surrounding area. In the preserve, this means they can usually be found on levees and spoil banks (Urbatsch, Ferguson, and Gunn-Zumo, 2007). These forests are dominated by communities of hardwood species such as live oak (*Quercus virginiana*), water oak (*Quercus nigra*) and American elm (*Ulmus americana*). Structurally, all three species are characterized by wide-spreading limbs, forming a broad, low, dense, symmetrical, round-topped crown. Live oaks typically range in height from 40 to 50 ft, water oaks from 50 to 80 ft, and American elms from 80 to 120 ft (Institute of Food and Agricultural Sciences, 2011).

In summary, the area from the western edge of the preserve to the Kenta Canal is covered by flat, homogenous marsh interrupted only by isolated patches of shrub-scrub marsh and bottomland hardwood forest along the spoil banks. East of the Kenta Canal, a gradient, mostly in tree size and density rather than species, exists between the marsh and baldcypress-tupelo swamp.

EAARL Vegetation Metrics

Composite Temporal Waveform Analysis

The data for this study were derived from an EAARL survey conducted at JELA during September 2006. At the nominal flying altitude of 300 m, a single EAARL laser pulse illuminates a small horizontal sampling area 20 cm in diameter. As a result, in a forest environment, the information content of the returned laser signal includes a very small portion of

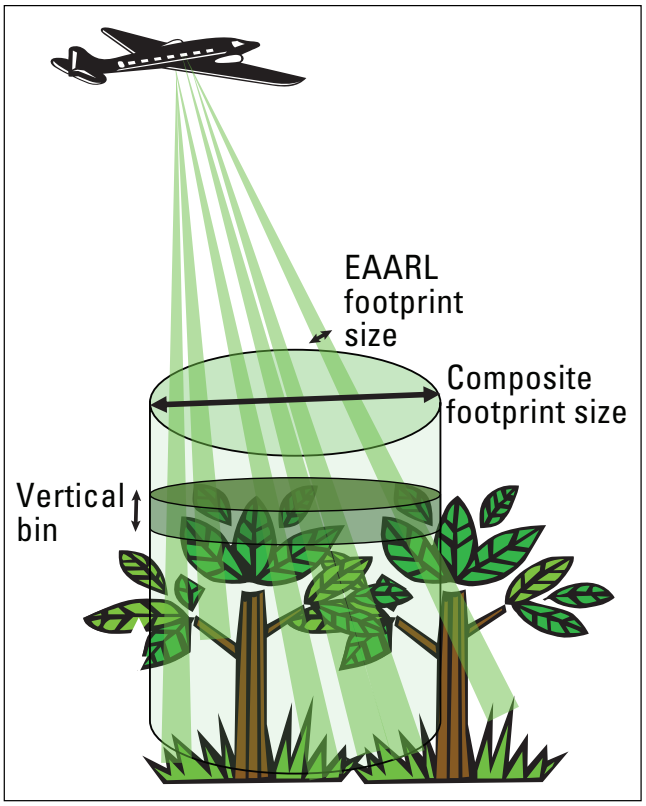


Figure 2. Schematic showing the composite footprint waveform principle (from Nayegandhi and others, 2006).
[EAARL is Experimental Advanced Airborne Research Lidar]

the canopy. This return waveform within a forest canopy may describe the vertical distribution starting from the side of a tree crown rather than the top or peak of the crown. Although a small footprint can improve the accuracy of the high-resolution digital elevation model (DEM) produced, a larger circle of illumination (footprint) increases the odds of encompassing the top of the tree crown. To describe the vertical structure of a vegetation canopy, several individual small-footprint laser pulses were combined to make a composite "large-footprint" waveform that defines a larger horizontal area (figs. 2 and 3).

The size of the composite footprint can be specified in the post-flight processing software and was set to 5 m × 5 m (Bonisteel and others, 2009). The composite footprint size depends on (1) the density of the lidar data (for this survey, the EAARL provided one laser pulse every 2-3 m^2), (2) the nature of the forested terrain (dense forested canopies are difficult to penetrate and require a relatively large composite footprint to describe the complete vertical structure), and (3) the desired horizontal resolution of the end product. A 5-m footprint size was determined to be optimal after testing several sizes and in consultation with NPS personnel. A 10-m footprint produced a result that was too coarse to be useful to park managers, whereas a 3-m footprint resulted in too few data points per grid-cell, resulting in holes in the datasets. The vertical-sampling resolution (or vertical bin) of the composite waveform was set to 50 cm. Within each vertical bin, the amplitude backscatter for all the individual waveforms constituting the composite waveform were averaged as follows:

$$\beta^{comp} = \frac{\sum_{i=1}^{n} \beta_i^{ind}}{n} \quad (1)$$

where β^{ind} is the backscatter count for each individual waveform i, n is the number of waveforms in the vertical bin, and β^{comp} is the resulting backscatter count for the composite waveform. The resulting composite waveform represents the vertical structure within a circular cone, similar to a single waveform return from a large-footprint lidar system. A small portion of the raw EAARL point cloud with the accompanying 5-m large-footprint grid (in red) overlain upon it is depicted in figure 3. Each blue point represents the center of a 20-cm footprint. The points enclosed within each grid cell are then averaged to compute the large-footprint waveform from which the CH, CRR and HOME metrics are derived. For this tile, the number of small-footprint waveforms averaged into a large-footprint waveform varies between 1 and 23, with a mean of 6.

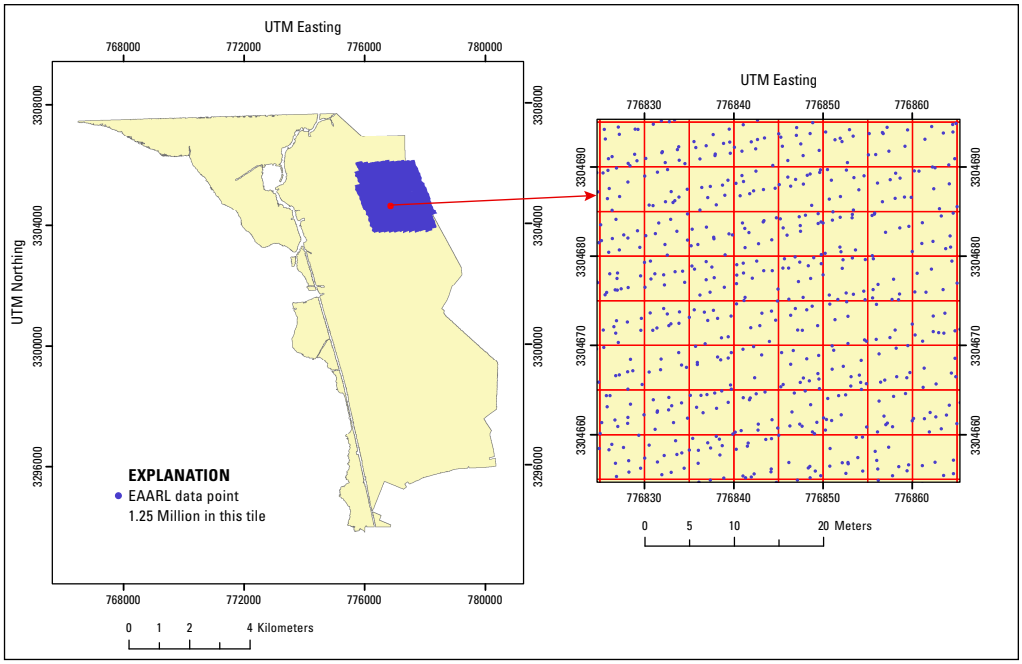


Figure 3. Map showing raw EAARL data points and 5-m large footprint grid.

Derivation of EAARL Metrics

BE elevations are determined from the range to the last peak in the *individual* small-footprint waveforms. A

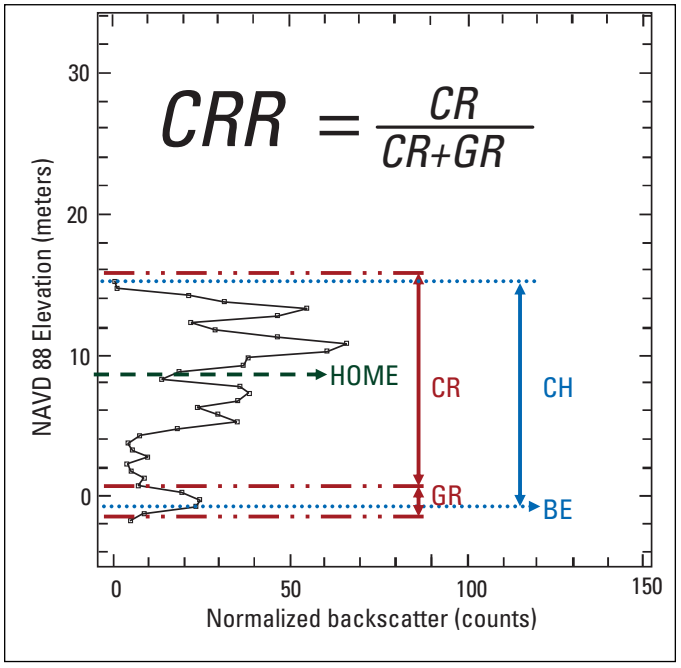


Figure 4. Graph showing vegetation metrics derived from a composite waveform.(BE–Bare Earth, CH–Canopy Height, CRR–Canopy Reflection Ratio, CR–Canopy Reflection, HOME–Height of Median Energy, GR–Ground Reflection).

"trailing-edge" algorithm was used to detect the range to the last peak (Bonisteel and others, 2009). This algorithm searches for the location at which the waveform changes direction for the last time along its trailing edge (Wagner and others, 2004) (fig. 4). The BE elevations are filtered with an iterative random consensus filter (RCF) and triangulated using Delaunay Triangulation (Shewchuk, 1996) to create a triangulated irregular network (TIN). These methods are explained in Nayegandhi, Brock, and Wright (2009). The TIN is then gridded to produce the 5-m × 5-m raster image that composes the BE metric (fig. 5).

The composite-footprint waveforms are used to derive three metrics: CH, CRR, and HOME. The definitions of these metrics and their components are illustrated in figure 4. The actual vegetation metrics derived from data collected during the 2006 EAARL survey of JELA are depicted in figures 5 through 8. CH is the distance from the first substantial peak in the waveform to the ground (fig. 6). The range to the first peak is detected at the first zero-crossing of the second derivative (Wagner and others, 2004) of the waveform. Accurate CH measurements are crucial for estimating various structural and (or) biophysical properties of forests, such as crown volume, basal area, and aboveground biomass (AGBM) (Hyyppä and others, 2001). The accuracy of the CH and BE metrics of EAARL has been established as being within 1-2 m for CH and within 20 cm for BE (Nayegandhi and others, 2006; Nayegandhi, Brock, and Wright, 2009). Canopy reflection (CR) is the sum of the portion of the waveform return reflected from all surfaces within the canopy. Likewise, ground reflection (GR) is the sum of the portion of the waveform return reflected off the ground. The CRR (fig. 7) is a relative measure of canopy cover and is defined as the ratio of CR to the total reflected energy of the ground and canopy:

$$CRR = \frac{CR}{CR + GR} \quad (2)$$

Independent knowledge about the average reflectance of the canopy and ground surfaces within the footprint is necessary to convert CRR to an absolute measure of canopy cover (Harding and others, 2001). HOME is the median height of the energy in the waveform above the ground, representing the elevation at which half the energy from the canopy occurs below and half above (fig. 8). This metric is predicted to be sensitive to changes in both the vertical arrangement of the canopy and the degree of canopy openness (including tree density). HOME has been found to be a good predictor of biomass and structural attributes in tropical forests (Drake and others, 2002). Applied in a similar manner to coastal vegetation communities, HOME could be used to look at structural changes across environmental gradients and, perhaps, to help assess damage to stands from storms and parasite infestations. CRR and HOME have been found to correlate well with commonly applied ground-based metrics in large-footprint lidars (Drake and others, 2002; Harding and others, 2001).

The Airborne Lidar Processing System (ALPS) software was used to develop the metrics just described. ALPS was developed in an open-source programming environment on a Linux platform. It combines the laser return backscatter digitized at 1-nanosecond intervals with aircraft positioning information and permits the exploration and processing of the EAARL data in either an interactive or batch processing mode.

Methodology

Five-meter-resolution grids, one pixel per composite waveform, of the four metrics (BE, CH, CRR and HOME) were used as input data for an unsupervised classification procedure executed within the open-source statistical software environment R (Institute for Statistics and Mathematics, 2009). Because of computing limitations, the data for each input and each output were composed of 34 2-km × 2-km tiles rather than one mosaic.

Clustering

Clustering involves finding similarities in a dataset and grouping the similar objects together. Thus, a cluster contains similar data that are dissimilar to data belonging to other groups. The similarity criterion or rule is usually either a minimizing objective function (for example, different types of geometric distance), or a descriptive concept. The classification method depends on the type of data, volume of data, and desired outcome (fig. 9).

For vegetation delineation, only exclusive clustering is appropriate because one "object" cannot belong to two different vegetation classes, which are usually defined as being mutually exclusive. In a supervised classification scheme,

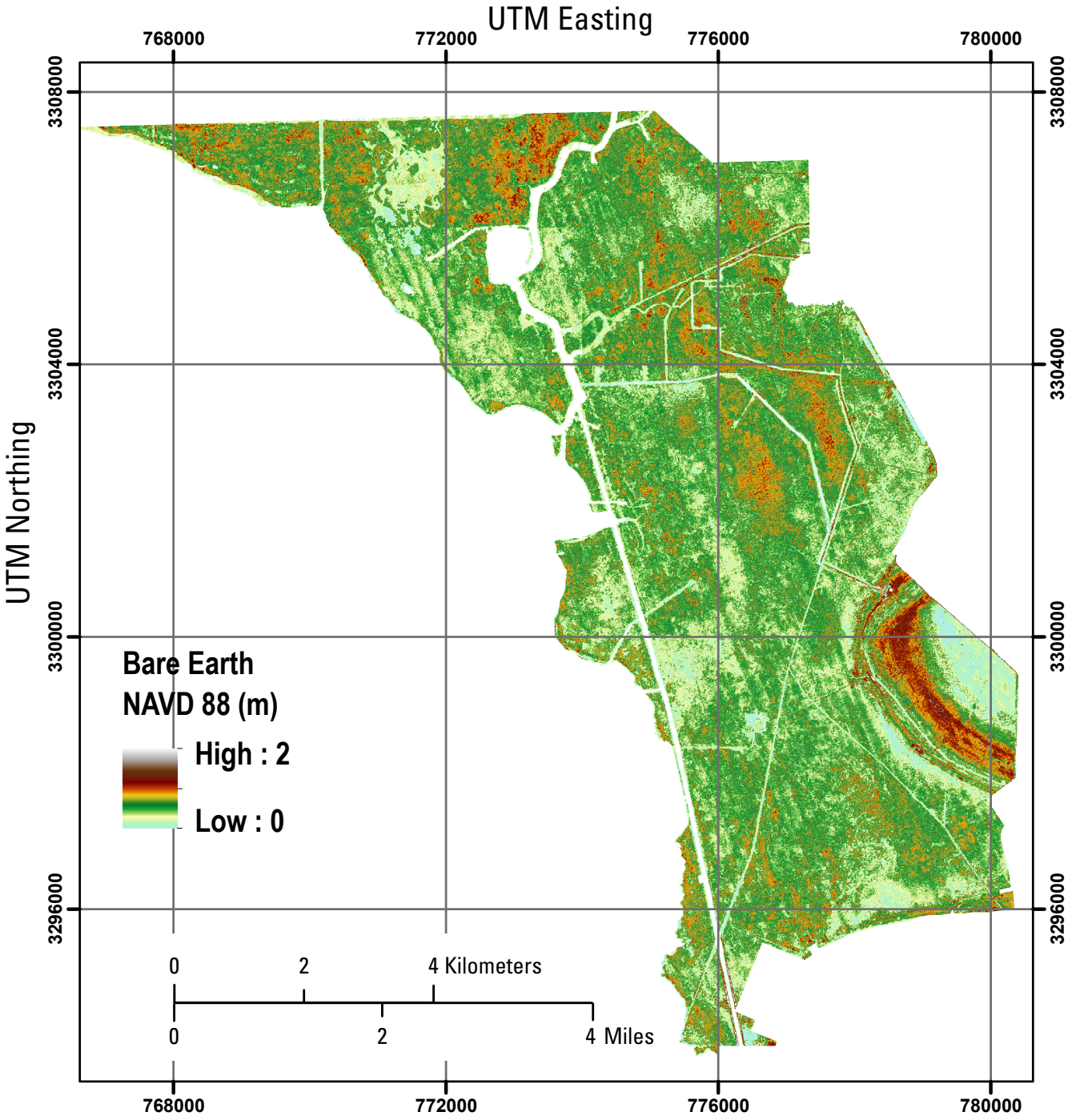


Figure 5. Map showing Bare Earth Digital Elevation Model (DEM) for JELA.

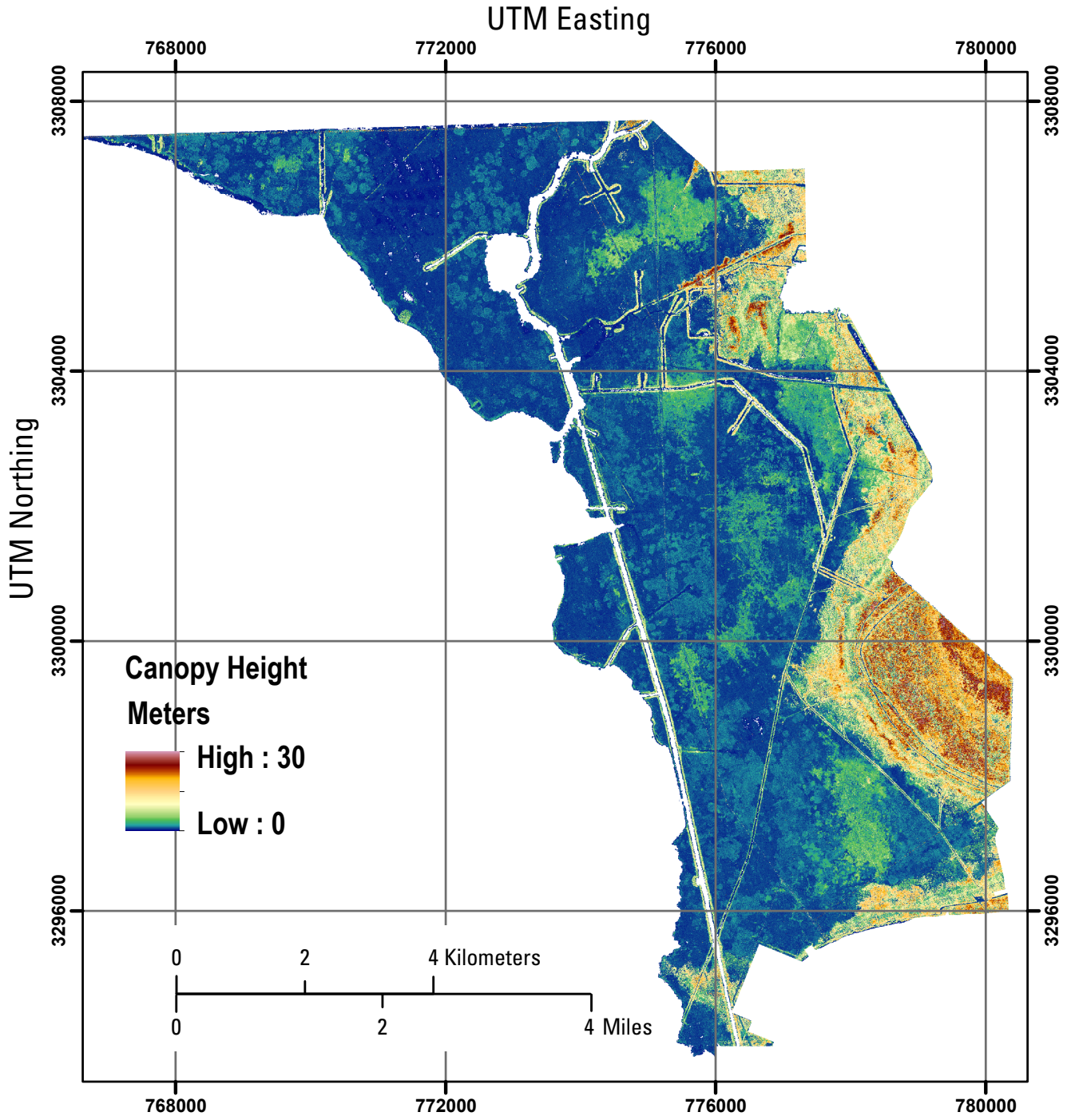


Figure 6. Map showing Canopy Height vegetation metric for JELA.

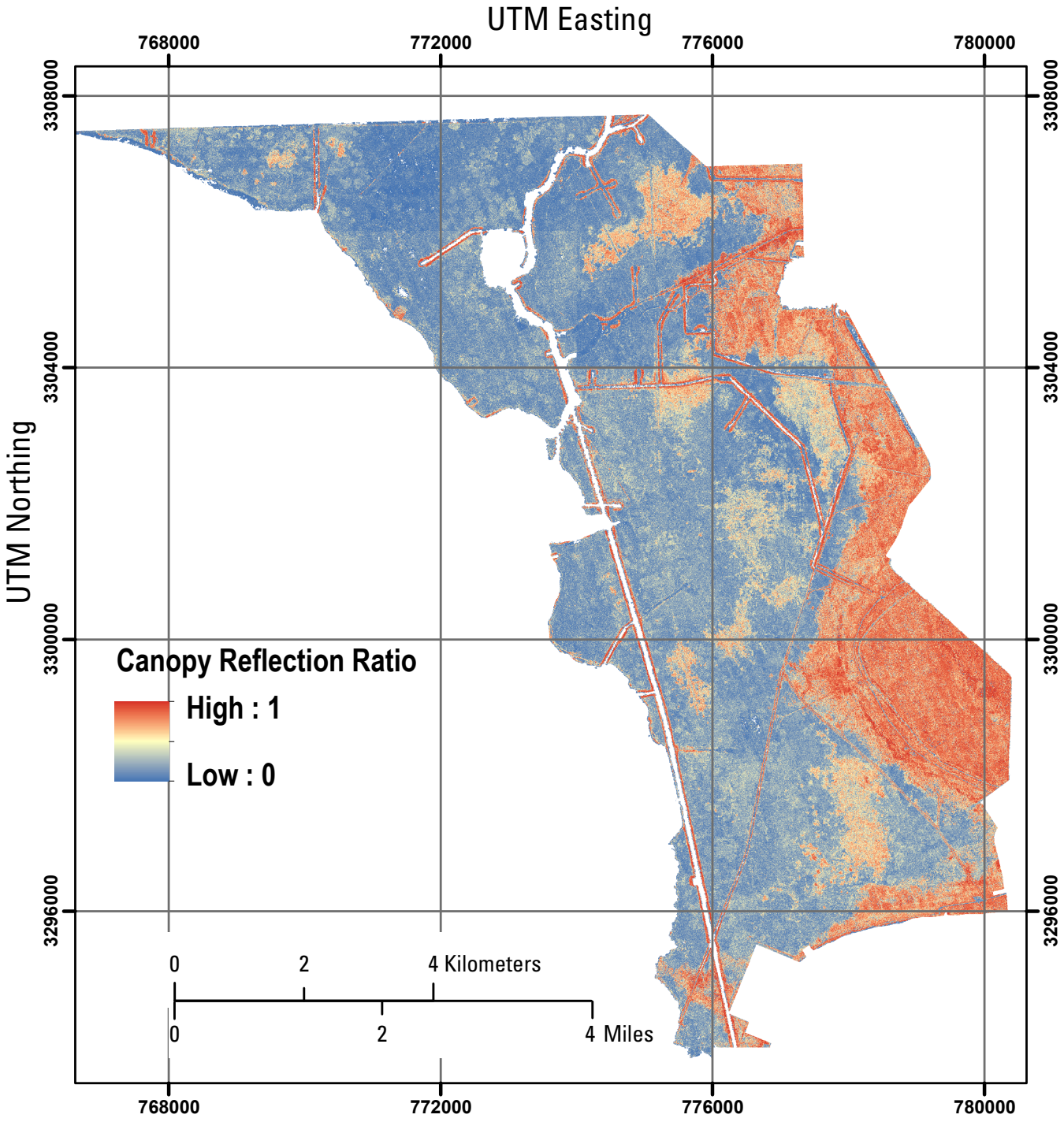


Figure 7. Map showing Canopy Reflection Ratio vegetation metric for JELA.

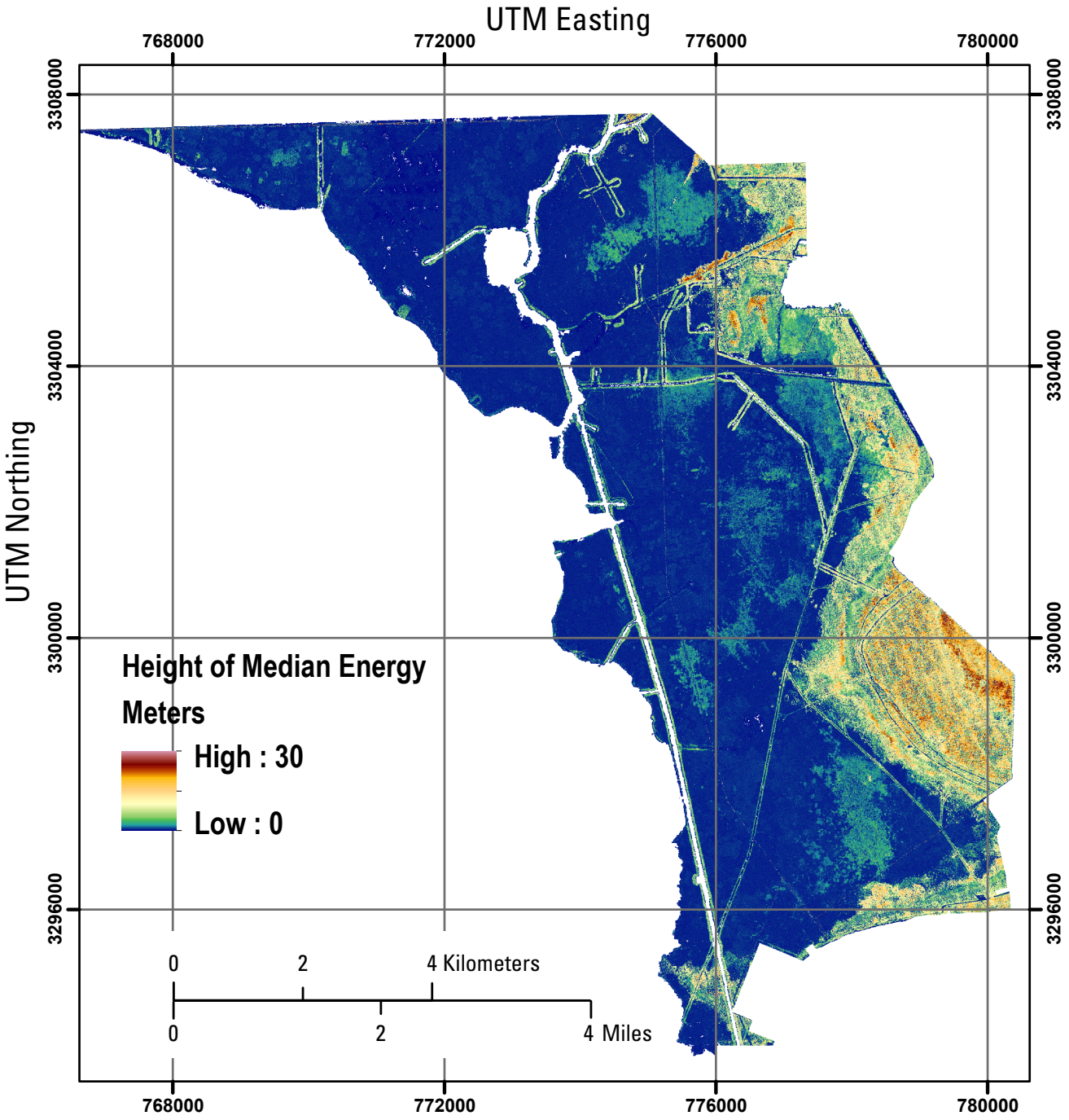


Figure 8. Map showing Height of Median Energy vegetation metric for JELA .

samples of points with known locations and classifications are manually tagged, and the clustering algorithm then classifies the remaining points based on their similarity to the tagged points. During the design phase of the project, no ground truth data for structural complexity existed for JELA. The use of a supervised (extrinsic) classification method was therefore precluded. Of the two types of intrinsic classification schemes available, partitioning methods are more compute-intensive but have the advantage of relocating data points as the classification process develops. This means that an incorrect decision (class assignment) made earlier in the classification process will be corrected as the decision criteria are refined. This very desirable property prompted the selection of an exclusive, intrinsic, partitioning classification method as highlighted in figure 9.

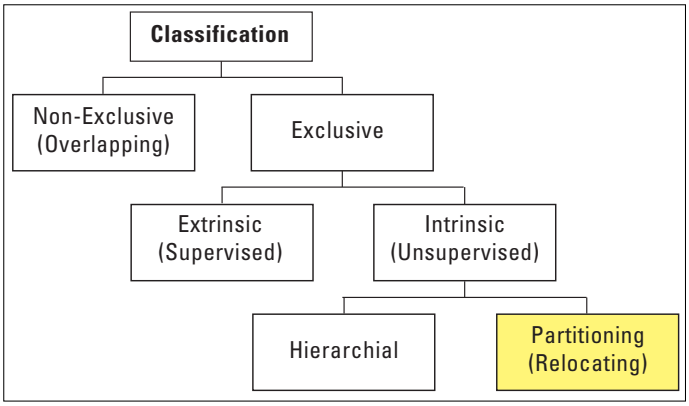


Figure 9. Diagram showing types of classifications (after Miranda, 1999). Schema used in this study is highlighted.

Data volume also restricts the choice of available clustering algorithms, as the EAARL metrics datasets tend to be very large, ranging from hundreds of thousands to several million data points. Using a 5-m × 5-m footprint results in 40,000 footprints per square kilometer. Because the Barataria Preserve is approximately 80 square km, each corresponding dataset contains more than 3 million pixels.

The clustering method selected for this analysis is the Clustering Large Applications (CLARA) algorithm described in Kaufman and Rousseeuw (1986, 2005). The CLARA algorithm is based on the k-medoid methods (Kaufman and Rousseeuw, 1986, 2005), but modified to efficiently handle large volumes of data. The k-medoid is the most representative data point within the k^{th} cluster (where k is the number of clusters). After finding the medoids, k clusters are created by assigning each observation in the dataset to the nearest medoid in such a way as to minimize the sum of the dissimilarities of the observations to their assigned representative object (Kaufman and Rousseeuw, 2005). Dissimilarity is measured by the Euclidian distance, which is the *square root of the sum of the squared differences*, between coordinates of a pair of objects in an n-dimensional space. The k-medoid clustering algorithms have several advantages, the most important being their robustness against outliers in the data because the choice of the medoids depends on the location of the majority of the points inside each cluster and is independent of the attribute type (Berkhin, 2002). The algorithm is implemented in the cluster package for R (Maechler and others, 2005).

After the initial classification is obtained, all classes are reordered along a median height gradient, with Class 1 having the lowest median height vegetation and Class k the highest median height vegetation. Also, in this case, Class 1 will have the simplest vegetation structure, and Class k a more complex one. Because the classes were reordered by median height, a shift from class n to class $n+2$ indicates a shift in vegetation height. However, this is not necessarily the case when interpreting change analysis results because the classes are not exclusively determined by vegetation height. The shift in class ranking can also be the result of changes in CRR, HOME, or both.

Unsupervised Classification

The metrics tiles were read into R as GeoTIFFs, filtered to remove "no-data" values, and saved as ASCII text files, one per tile, with the four metrics values recorded for each pair of XY coordinates in the tile. A principal component analysis (PCA) of the remaining data was performed, and each principal component was multiplied by its respective proportion of variance explained. A high performance k-medoids algorithm (Kaufman and Rousseeuw, 2005; Struyf, Hubert, and Rousseeuw, 1997) was used to conduct an unsupervised classification with eight classes. The results were then reclassified such that Class 1 data had the lowest median canopy height and Class 8 had the highest median canopy height. The reclassified result was then transformed back to 2-km × 2-km GeoTIFF raster images, which can be opened in any software capable of reading projected GeoTIFF files such as Global Mapper, ENVI, ERDAS or ArcGIS. The complete procedure is detailed in the methodology flowchart (fig. 10).

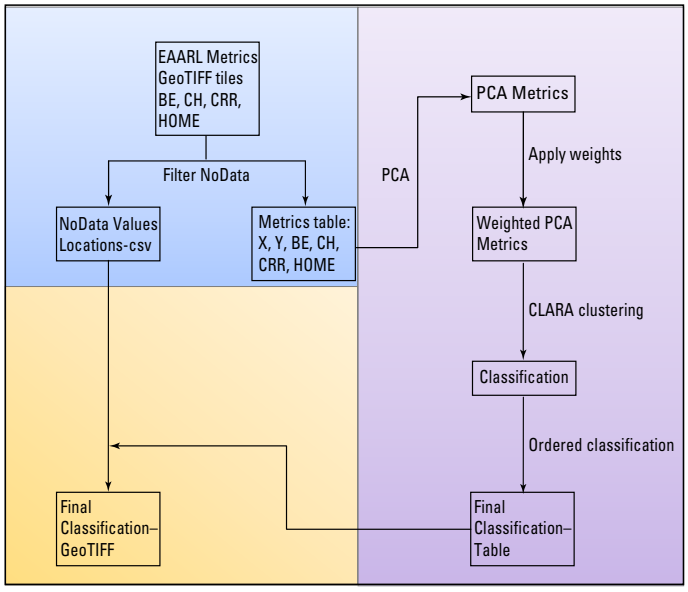


Figure 10. Methodology flowchart.

Results and Discussion

To establish which of the vegetation metrics were most influential on the classification, we examined the clusters' spatial distribution. Analysis of figures 11 and 12, along with table 1, indicates that CH and HOME influence the classification the most, followed by CRR and, finally, BE.

Influence of Principal Components and Vegetation Metrics on Classification

In figures 11 and 12, the x-axis variable is the named box in the same column, whereas the y-axis variable is the named box in the same row. For example, in figure 11, the plot in the first row of the second column is a scatterplot of BE as a function of CH for all the points in the JELA dataset, color-coded by class. Similarly, the plot in the second row, last column is a scatter plot of CH as a function of HOME.

Table 1. Correlation matrix of vegetation metrics.

Vegetation metric	BE	CH	CRR	HOME
BE	1.00	0.00	0.00	-0.01
CH	0.00	1.00	0.79	0.96
CRR	0.00	0.79	1.00	0.73
HOME	-0.01	0.96	0.73	1.00

Upon visual inspection, it is apparent that although discernible trends exist, discriminating all classes based on the metrics alone is difficult or impossible because of mixing of points in the feature space. This mixing is a result of the high correlation (greater than 0.70) between CH, CRR, and HOME, and suggests that an orthogonal transformation of the input metrics would produce better class separation. The orthogonal transform selected, PCA, was carried out and the classification

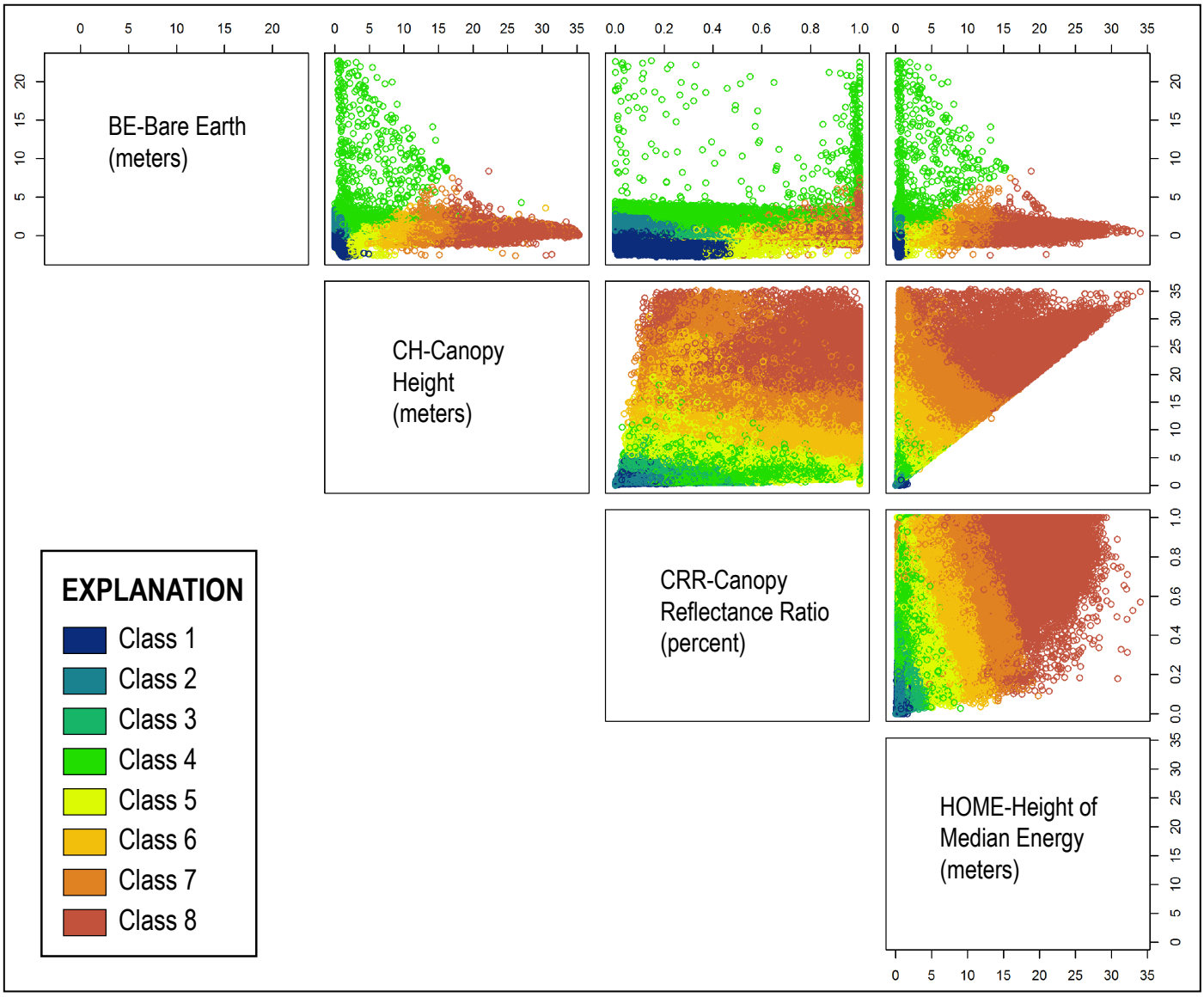


Figure 11. Scatterplots of metrics color-coded by class.

was performed on the principal components, which are, by definition, uncorrelated. Correlations between metrics are displayed in table 1.

The improved class separation is evident in the figure 12 scatterplots between the principal components of the JELA dataset, again color-coded by class. Here, c1 represents component 1, c2 component 2, and so forth; the component axes are dimensionless. The underlying principle of PCA is that the first component contains the highest percentage of the variance in the input dataset (in this case the metrics) and is therefore the most significant, the second component contains the next highest percentage of variance, and so forth; the last component is typically mostly noise. The percentage of variance is a measurable quantity that can be related to the inputs. For this dataset, c1 explained 66.4 percent of the variance, c2 explained 25 percent, c3 explained 7.7 percent, and c4 explained less than 1 percent. The principal components, weighted by their percentage of variance explained, were then input into the classification algorithm.

Table 2. Correlation matrix of vegetation metrics and principal components.

Vegetation metric	Principal component			
	C1	C2	C3	C4
BE	0.001	1.000	0.00	-0.001
CH	-0.976	0.00	0.165	0.140
CRR	-0.887	0.00	-0.462	-0.020
HOME	-0.958	0.00	0.259	-0.125

That component 1 is indeed the most important to the classification results is reflected in the clear stratification by class that is visible in the first row of scatterplots in figure 12. These plots indicate that classes 6, 7, and 8 (vegetation classes with median heights above 8.5 m) are almost exclusively determined by the first principal component (c1). Correlations between the metrics and their principal components are listed in table 2. Very strong correlations (greater than 95 percent)

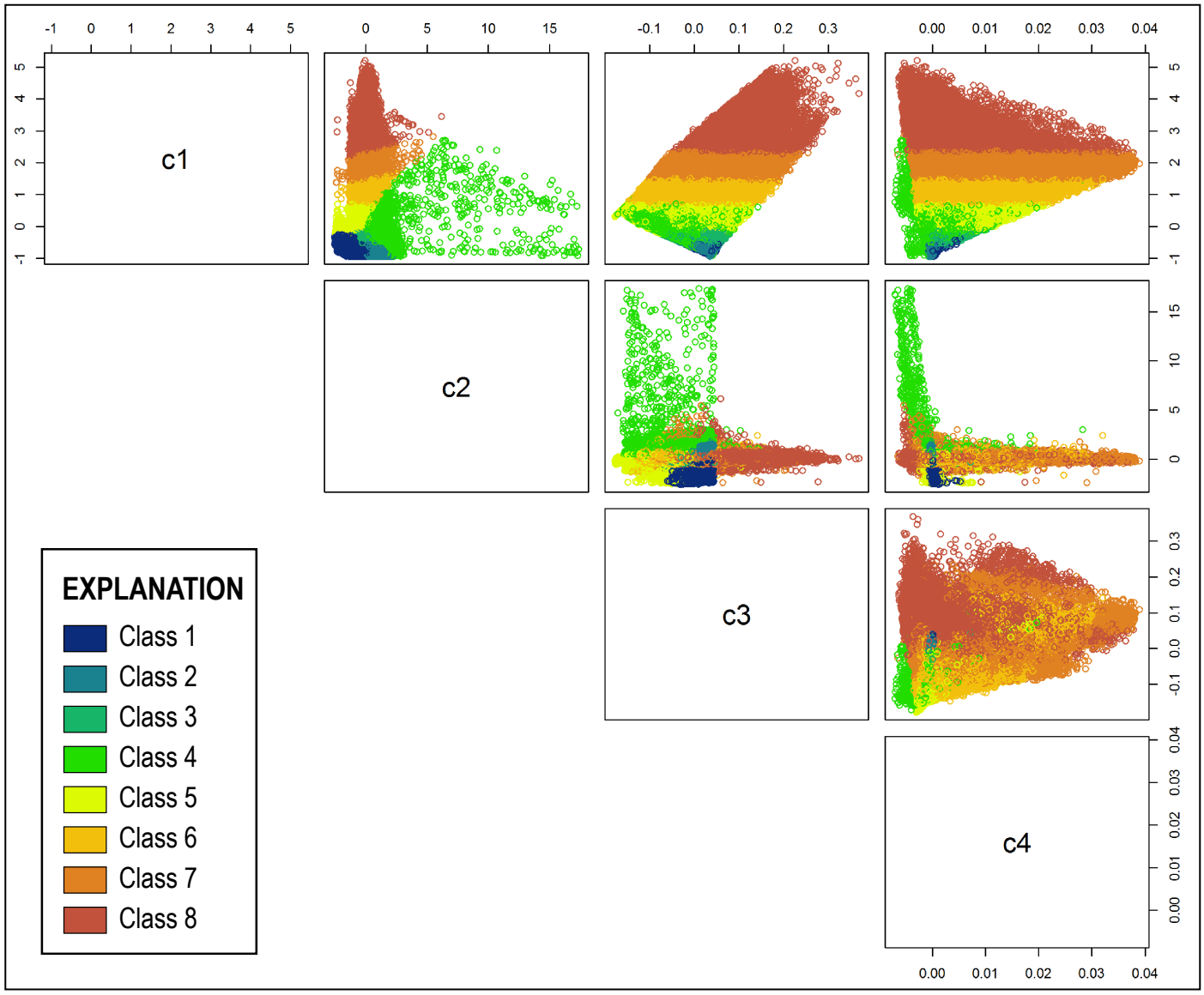


Figure 12. Scatterplot of principal components color-coded by class. Component axes are dimensionless.

between the first principal component and CH and HOME imply that these metrics are dominant in terms of class discrimination, especially those of taller vegetation. CRR also has a strong correlation to c1 (88 percent), suggesting that it is of tertiary importance to the classification behind the CH and HOME metrics. The c1–c2 scatterplot indicates that c2 is influential in discriminating classes 1 and 2, which represent water, bare ground, and vegetation having median heights below 0.86 m. The 100 percent correlation between c2 and ground elevation (BE) indicates that BE is therefore similarly influential on the discrimination of classes 1 and 2. Although the correlations of the input metrics to c3 are somewhat weak, c3 is most strongly correlated with CRR and, when used in combination with c1, presents a clear and unassailable discrimination of classes 6 to 8. C2 and c3 taken together supplement discrimination of classes 1 and 2. C4, although noisy, supports the discrimination of class 4, when used in conjunction with c2 (based on the c2–c4 scatterplot).

Ground Truth

Ninety-five NPS sampling locations (fig. 13) taken from a larger 2007 NPS study were used as ground truth sites to determine how well the lidar metrics characterize vegetation structure. At these locations in JELA, local ecologists sampled specific 10-m radius ground transects for species composition, canopy cover, and ground cover. For each sampling location, the average of each of the four metrics (BE, CH, CRR, and HOME) within a corresponding 10-m radius transect was calculated and plotted against percent canopy cover. Ground cover types that do not contribute to canopy cover (water, bare ground, coarse woody debris and grasses) were excluded from this analysis. Both classical (Pearson) and robust correlation coefficients were calculated between each metric and percent canopy cover. The classical correlation coefficient is a function of the mean and standard deviation of the dataset and thus is very susceptible to the influence of outliers in the data. In contrast, the robust correlation coefficient is designed specifically to not be unduly affected by outliers and other small departures from model assumptions. The robust correlation algorithm selected (Rousseeuw and Leroy, 1987) uses the minimum covariance determinant (MCD) estimator in the search for the optimal subset of data points, which is the subset with the highest (classical) correlation. The robust correlation coefficient reveals the background correlation of the majority of the data, thereby mitigating scattered deviations. The minimum size of the subset was set at 50 percent +1, and any points not included in the resulting subset were considered outliers.

The results of the correlation analysis are shown in figure 14 and indicate strong positive correlations between the CH, CRR, and HOME metrics with percent canopy cover and a

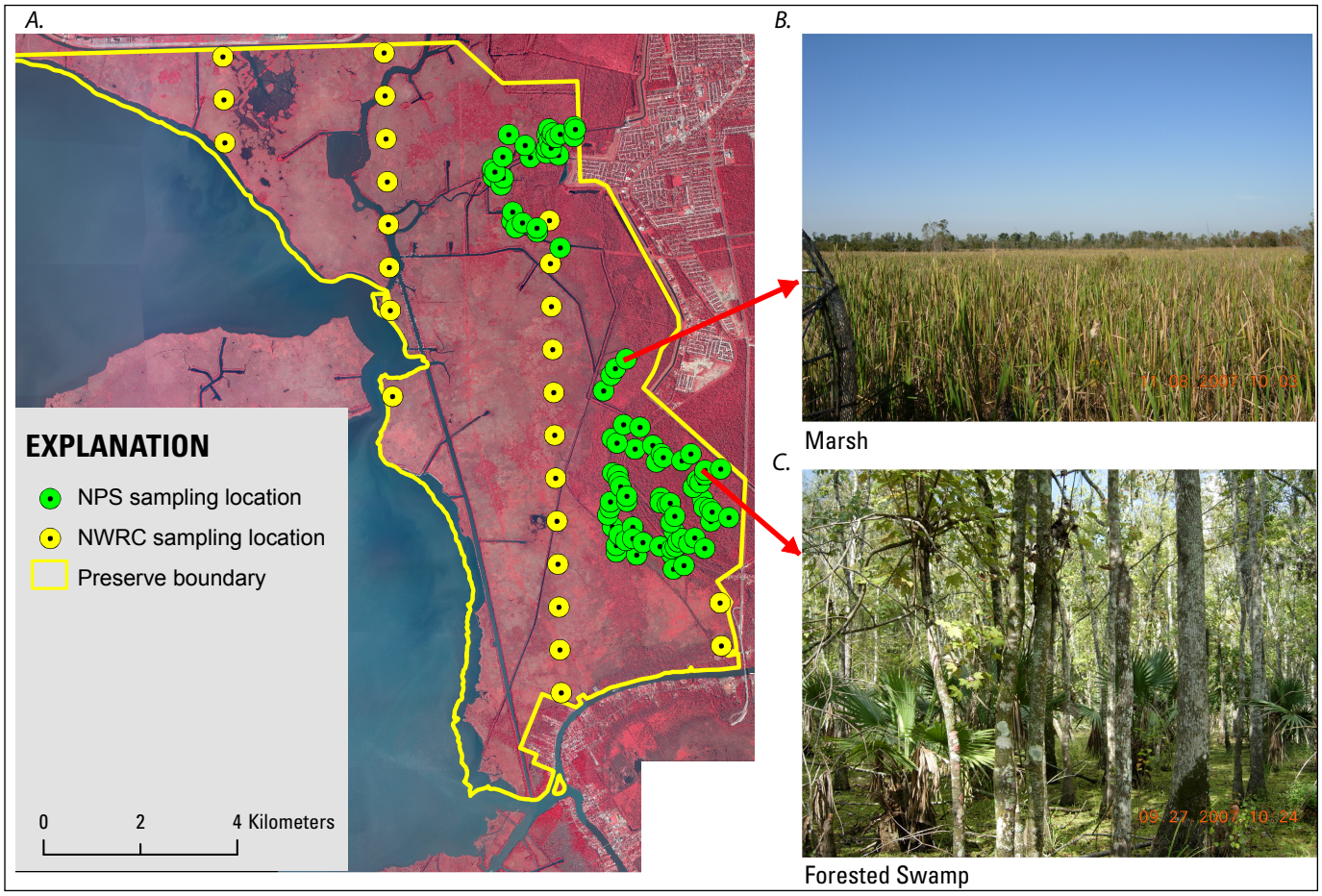


Figure 13. *(A)* Map showing sampling locations, and field photographs showing *(B)* marsh and *(C)* forested swamp.

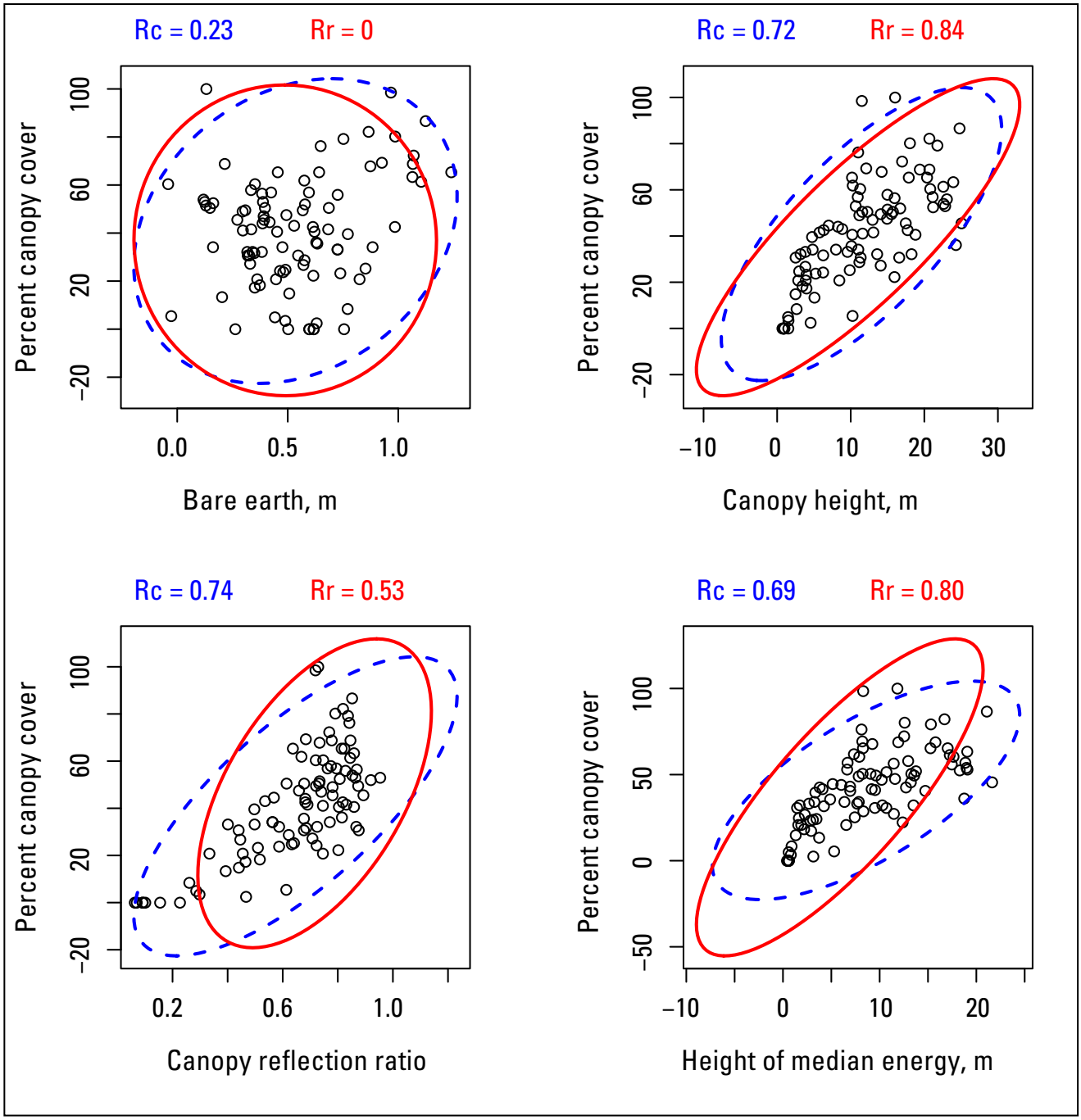


Figure 14. Graphs showing distribution of EAARL metrics with percent canopy cover. Rc represents classic correlation coefficient; Rr represents robust correlation coefficient.

Table 3. Correlations between EAARL metrics and percent canopy cover.

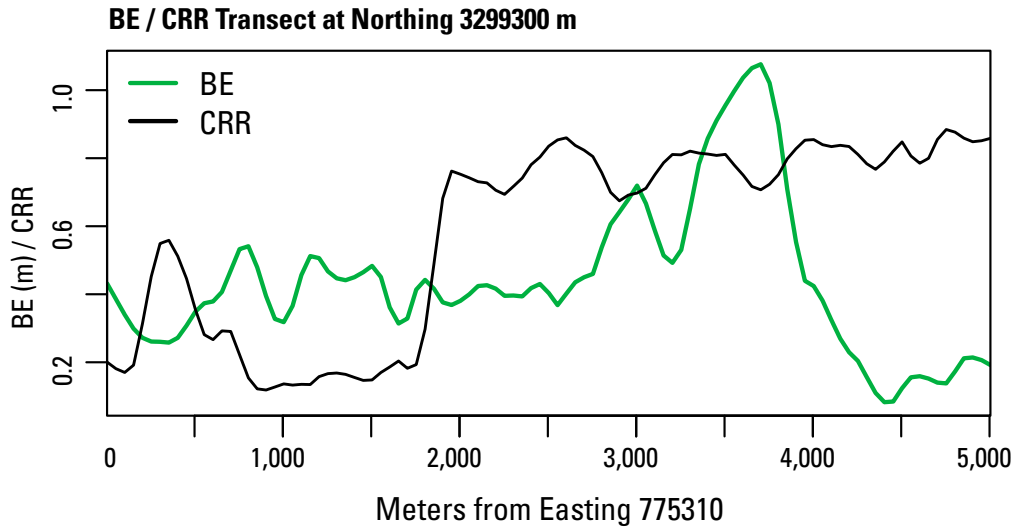
Metric	Classic correlation	p-value	Robust correlation	p-value
BE	0.23	0.029	-0.00	0.979
CH	0.72	4.44×10^{-16}	0.84	1.59×10^{-13}
CRR	0.74	$<2.2 \times 10^{-16}$	0.53	0.0001
HOME	0.69	1.78×10^{-14}	0.80	6.93×10^{-12}

[EAARL is Experimental Advanced Airborne Research Lidar]

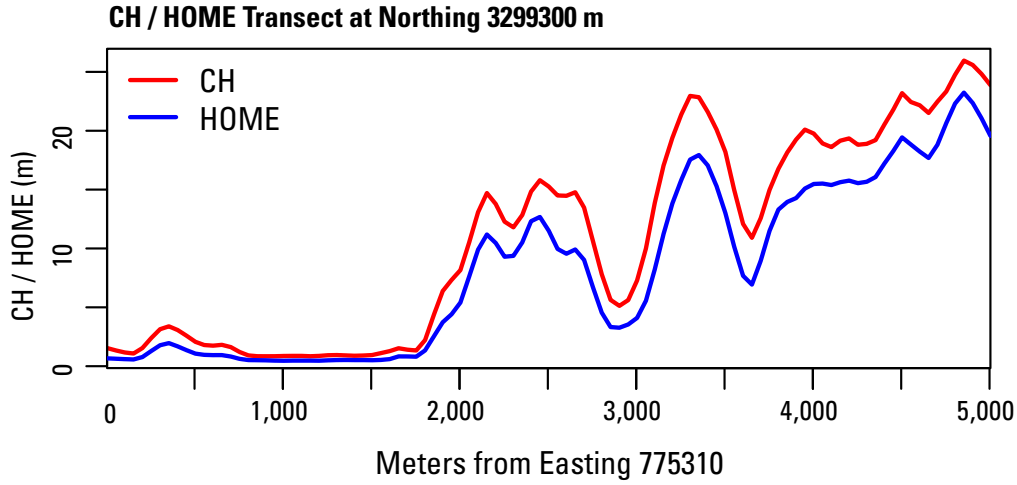
weak positive correlation between BE and percent canopy cover. High, statistically significant (table 3) correlation coefficients, along with robust statistics that are greater than the classical statistics obtained for CH and HOME, suggest that these metrics accurately reflect canopy cover, and thus, are suitable inputs for classifying vegetation assemblages in JELA. CRR also exhibits a strong classical correlation (74 percent) with percent canopy cover, but the lower robust statistic (53 percent) indicates that this metric is being affected by outliers.

Previous research (Franklin, Connery and Williams, 1994; Denslow and Battaglia, 2002; Miller and Franklin,

A.



B.



C.

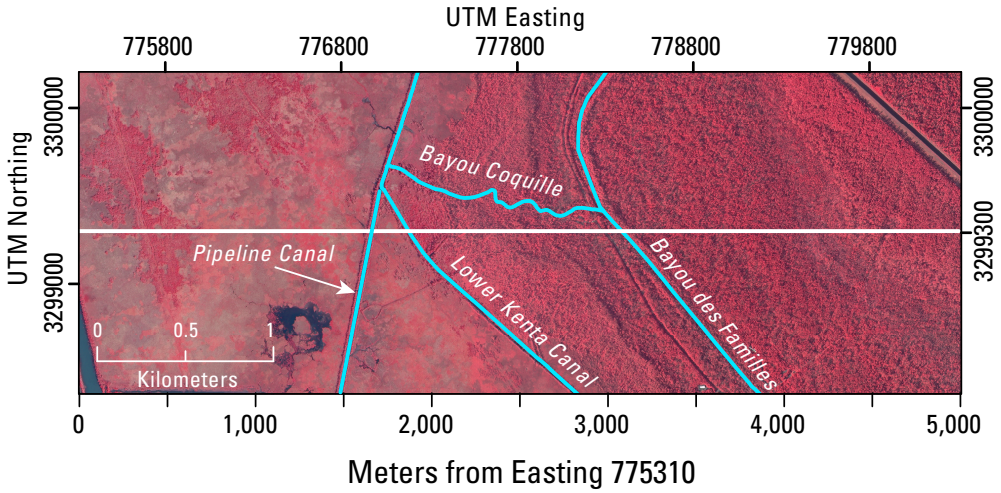


Figure 15. *(A, B)* Transect profiles of metrics with *(C)* corresponding color-infrared (CIR) imagery. X-axis distances are in meters from the beginning of the transect line.

2002; Salovaara and others, 2005; Sesnie and others, 2008) has frequently found vegetation type to be correlated with elevation. Most of the research studies, however, with the exception of the Denslow and Battaglia (2002) study, take place in regions with elevation gradients ranging from 80 to more than 3,000 m. At that scale, elevation definitely influences vegetation type. However, we contend that the small range of elevation in JELA (approximately 2 m) limits elevation effects on vegetation structure. This is illustrated by the low correlation between BE and percent canopy cover (fig. 14). The Denslow and Battaglia (2002) study, which also was conducted in JELA, mapped vegetation by species and structure and found topographic position to be a strong predictor of vegetation structure and species composition. The key difference in this analysis is that our variable of interest is percent canopy cover rather than species or number/size of stems. Percent canopy cover can be thought of as the percentage of plot area covered by tree crowns. Given this definition, a homogenous spartina marsh or mowed meadow, for example, would have a percent canopy cover of zero. Additionally, although the species present are very different, stands of both the bottomland hardwoods populating the top of the levee as well as the cypress populating the swamp have the full range of percent canopy cover: low for ground truth points taken in open spaces, intermediate at points which include canopy gaps, and high otherwise.

Transect Profiles of EAARL metrics

A plot of each metric along a 5,000-m east-west transect located at Northing 3299300 m (UTM 15/NAD 83) is shown in figure 15 along with the corresponding transect line on the CIR image; the data were smoothed before plotting. Although smoothing sometimes blurs nearby features together, it reduces noise in the signal, allowing only the most prominent features to remain. Several of these features, visible in the CIR image, are also discernible in the profiles of the metrics. X-axis distances on the transect graphs are in meters from the beginning (left end) of the transect line. From left to right, a strong response from the CRR plot 400 m from the beginning of the transect confirms the presence of a patch of scrub-shrub marsh. From there, the flotant marsh grows mostly uninterrupted until the Pipeline Canal at around 1,600 m. The Pipeline Canal is discernible as a clearly defined dip in the BE profile. Between the Pipeline Canal and the Kenta Canal, the transition to forest begins and all three canopy metrics begin trending upwards accordingly. A strong peak in the BE profile at 3,000 m, along with corresponding minima in CH and HOME, marks the location of State Highway 45/Barataria Boulevard. This is followed at 3,700 m by the highest peak in the BE profile, indicating the top of the eastern Bayou des Familles levee. Down the backslope of the levee and into the swamp, the elevation profile falls off rapidly while the CH and HOME metrics keep increasing. There are also several features noticeable on the profiles that cannot be seen at the 1:30,000 scale of the CIR image, such as thinning in the stands atop the levee.

Classification Results

Finally, the classification results were validated against 25 National Wetland Resource Center (NWRC) data points within the park boundary (fig. 13) from the well-known 2007 Louisiana Coastal Marsh–Vegetative Type Dataset (U.S. Geological Survey National Wetlands Research Center, 2008). This marsh vegetation inventory has been conducted once every 5 to 10 years since 1949, and involves biologists flying predefined transects above the marsh in a helicopter and identifying dominant plant species. The transects are spaced at 3-km intervals and the sampling points along each transect are spaced at approximately 800-m intervals. The survey of the section covering JELA occurred during late August, 2007. The results of the comparison were very favorable, with 20 of 20 sample locations classified as grasses and short vascular plants (*Typha L., Sagittaria lancifolia L., Polygonum L., Panicum L., Eleocharis R. Br., Spartina patens, Schoenoplectus californicus*) in the NWRC dataset occurring in Classes 1 to 3 in our classification, 2 of 2 sample locations classified as shrubs occurring in Class 3, and 2 of 2 sample locations classified as upland forest trees assigned to Classes 6 and 8. One data point in the NWRC dataset was unclassified, but inspection of CIR imagery shows it is on the edge of a spoil bank stand. This point is a member of Class 5, a reasonable assignment given that it lies on the transition from the spoil bank, which usually has tall, dense (Classes 6–8) vegetation, and the marsh, which is usually covered by Class 1 to 3 vegetation. Collectively, these results translate into a 100 percent correspondence despite the 1-year difference between the lidar and NWRC surveys.

Filtering

In order to reduce the high local variance observed in the JELA vegetation classification, the merged classification result was smoothed by applying a median filter with a 3 × 3 cell neighborhood in ArcGIS 9.3. This procedure yielded more homogeneous patches of vegetation (fig. 17), but concealed some of the park's smaller linear features (such as canals and roads). The median filter reduces local variation and removes noise in the data for a 3 × 3 cell neighborhood (15 m × 15 m) by replacing unique pixels with the median value of the pixels that surround them. Essentially, the median filter generalizes the data by removing high-frequency changes, leaving behind trends that are suitable for viewing at a larger scale. This filtering resulted in a classification map in which 68 percent of the classified area still was represented by Classes 1 to 3, 13 percent by Classes 4 to 5, and 19 percent by Classes 6 to 8. This near-identical distribution of classes indicates that filtering does not alter the profile of the classification, while having the benefit of allowing major trends to emerge. In practice, if the interest is in determining the likely type (for example, tall or short, open or dense) of cover at a given spot, then the unfiltered classification (fig. 16) is more appropriate. If the main objective, however, is to understand the spatial

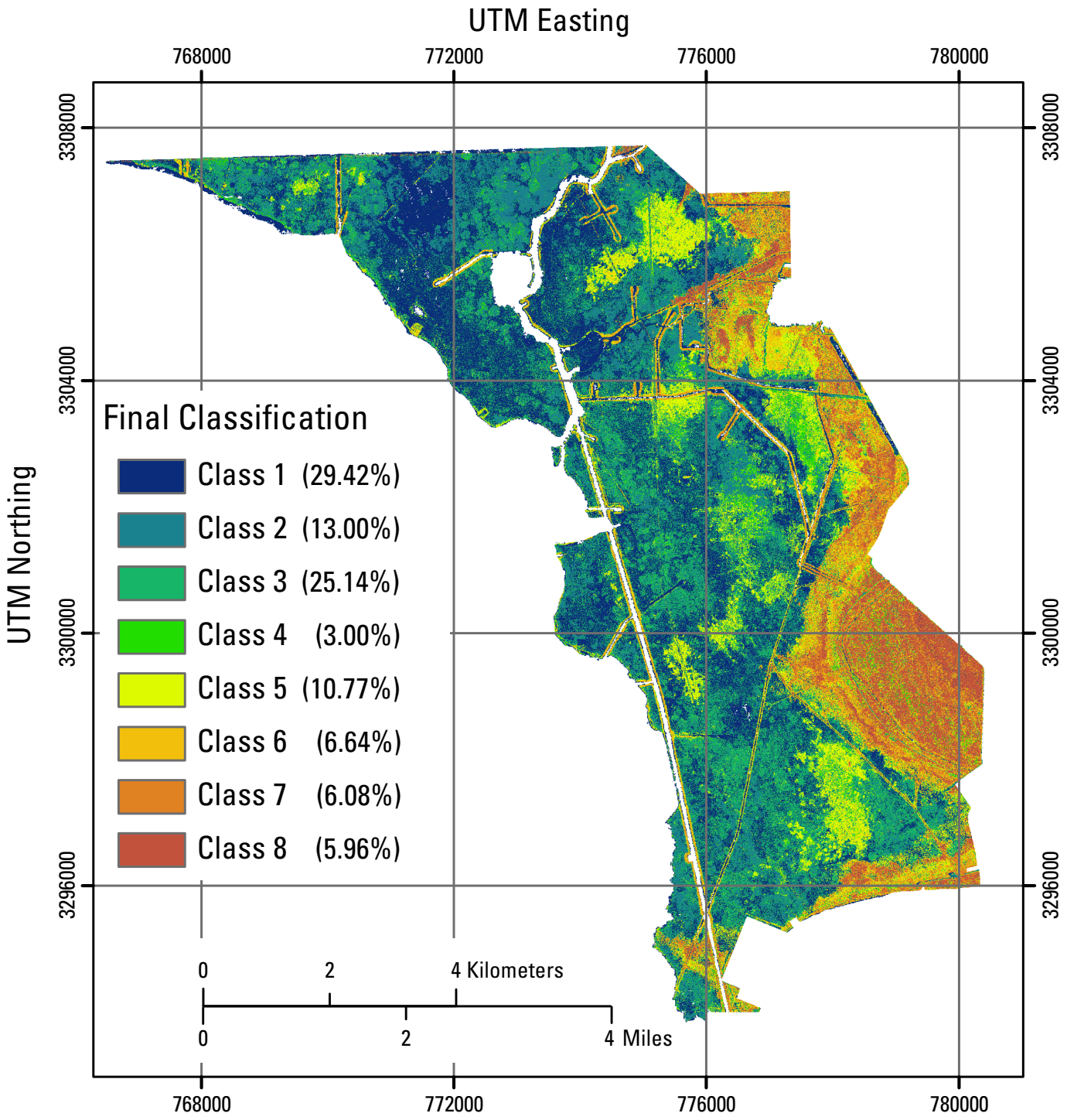


Figure 16. Map shoiwng final classification.

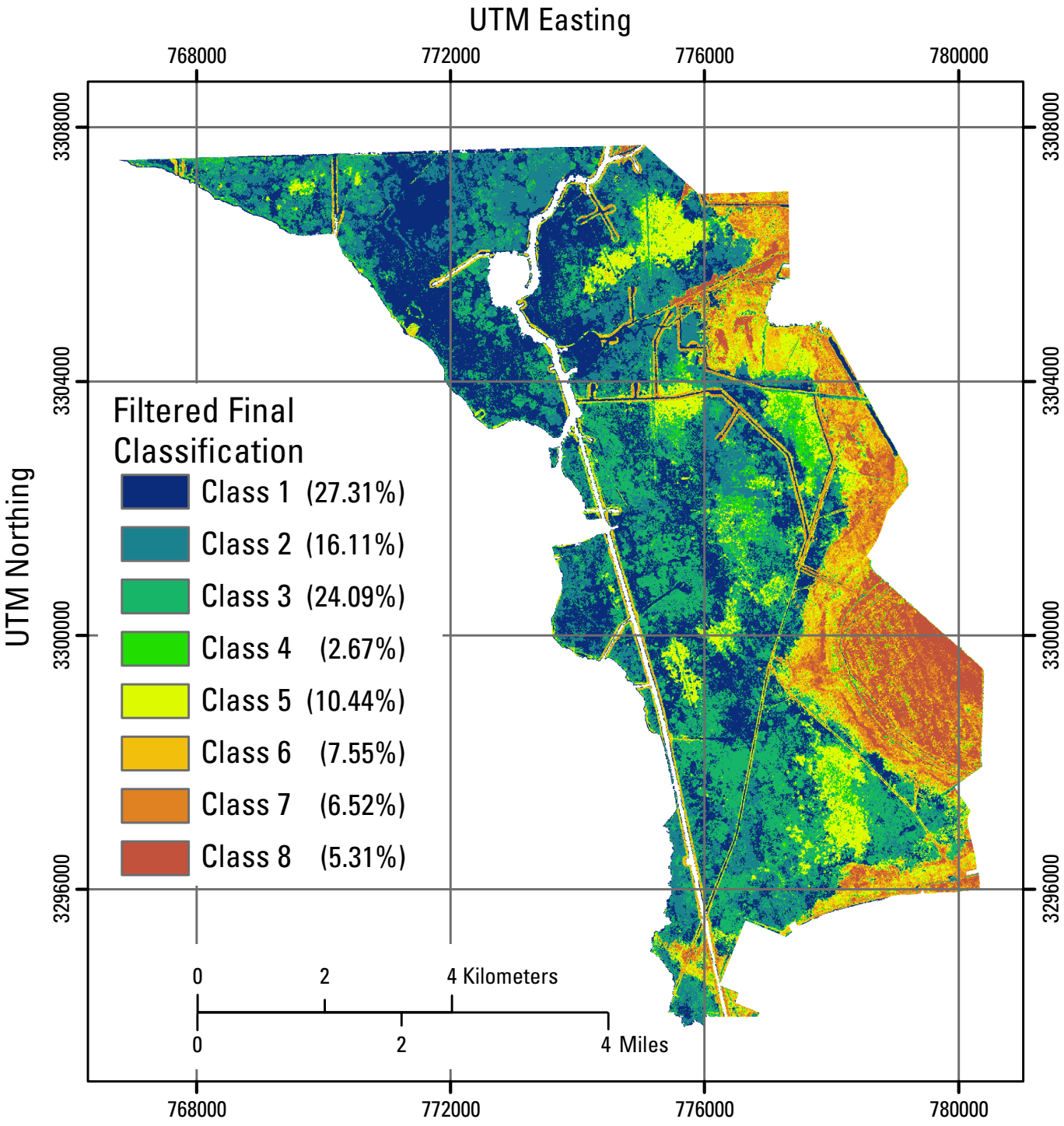


Figure 17. Map shoiwng filtered final classification.

distribution of broad swaths of different types of vegetation cover, then the filtered result (fig. 17) is more suitable. Both smoothed and unsmoothed classification results were provided to the NPS Gulf Coast Network.

Conclusions

This study evaluated the ability of the EAARL system for delineating coastal vegetation structural communities at JELA's Barataria Preserve using an unsupervised classification approach. The preserve includes distinct fresh marsh, flotant marsh, scrub-shrub marsh, bottomland hardwood forest, and forested swamp areas. In previous related publications on EAARL data, Nayegandhi and others (2006) and Nayegandhi, Brock, and Wright (2009) established the accuracy of the CH and BE metrics by concurrent field measurement of actual canopy heights and ground elevations, and found them to be accurate to within 1–2 m for CH and within 20 cm for BE. In this study, we established the relationship of CH, CRR and HOME relative to field measurements of percent canopy cover with an approximately 70 percent correlation between each canopy metric and the measured quantity. Principal components of the four metrics were then used as input to the CLARA algorithm—a high-performance unsupervised classification algorithm optimized for large datasets. Sixty-eight percent of the resulting eight-class classification resided in Classes 1–3, 14 percent in Classes 4–5, and 18 percent in Classes 6–8. The classes were arranged along an increasing median height gradient. Although each of the metrics has the potential to dominate the classification, for this dataset it was found that CH and HOME were most influential, followed by CRR and, finally, BE. The final classification map agrees closely with features visible on 1-m aerial imagery, as well as with the independent 2007 Louisiana Coastal Marsh–Vegetative Type Dataset (U.S. Geological Survey National Wetlands Research Center, 2008), despite the 1-year time lag between the EAARL survey and both datasets (NPS and NWRC) used for ground truth.

References Cited

Battaglia, L.L., Denslow, J.S., and Hargis, T.G., 2007, Does woody species establishment alter herbaceous community composition of freshwater floating marshes?: Journal of Coastal Research v. 23, no. 6, p. 1580–1587.

Berkhin, Pavel., 2002, Survey of clustering data mining techniques: San Jose, Calif., Accrue Software, accessed September 2009, at *http://www.ee.ucr.edu/~barth/EE242/clustering_survey.pdf*.

Bonisteel, J.M., Nayegandhi, Amar, Wright, C.W., Brock, J. C., and Nagle, D.B., 2009, Experimental Advanced Airborne Research Lidar (EAARL) data processing manual: U.S. Geological Survey Open-File Report 2009-1078, 38 p.

Denslow, J.S., and Battaglia, L.L., 2002, Stand composition and structure across a changing hydrologic gradient: Jean Lafitte National Park, Louisiana, USA: Wetlands, v. 22, p. 738–752.

Drake, J.B., Dubayah, R.O., Knox, R.G., Clark, D.B., and Blair, J.B., 2002, Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest: Remote Sensing of Environment, v. 81, p. 378–392.

Franklin, S.E., Connery, D.R., and Williams, J.A., 1994, Classification of alpine vegetation using Landsat Thematic Mapper, SPOT HRV and DEM data: Canadian Journal of Remote Sensing, v. 20, p. 49–58.

Harding, D.J., Lefsky, M.A., Parker, G.G., and Blair, J.B., 2001, Laser altimeter canopy height profiles: Methods and validation for closed-canopy, broadleaf forests: Remote Sensing of Environment, v. 76, p. 283–297.

Homer, Colin, Huang, Chengquan, Yang, Limin, Wylie, Bruce, and Coan, Michael, 2004, Development of a 2001 National Landcover Database for the United States: Photogrammetric Engineering and Remote Sensing, v. 70, no. 7, p. 829–840.

Hyyppä, Juha, Kelle, Olavi, Lehikoinen, Mikko, and Inkinen, Mikko, 2001, A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners: IEEE Transactions on Geoscience and Remote Sensing, v. 39, p. 969–975.

Institute of Food and Agricultural Sciences, 2011, Florida Forest Stewardship: University of Florida, accessed November 2011, at *http://www.sfrc.ufl.edu/Extension/florida_forestry_information/forest_resources/*.

Institute for Statistics and Mathematics, 2009, The Comprehensive R Archive Network: WU Wirtschaftsuniversität Wien, Vienna University of Economics and Business, accessed September 2009, at *http://cran.r-project.org*.

Kaufman, Leonard, and Rousseeuw, P.J., 1986, Clustering large data sets *in* Pattern recognition in practice II, Gelsema E.S., and Kanal L.N., eds.: Elsevier/North-Holland, p. 425–437.

Kaufman, Leonard, and Rousseeuw, P.J., 2005, Finding groups in data: An introduction to cluster analysis: Hoboken, N.J., Wiley, 342 p.

Maechler, Martin, Rousseeuw, Peter, Struyf, Anja, Hubert, Mia, and Hornik, Kurt, 2005, Cluster analysis basics and extensions: cluster R package version 1.11.11, accessed September 2009 at *http://www.cran.r-project.org/web/packages/cluster/*.

Miller, Jennifer, and Franklin, Janet, 2002, Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial

dependence: Ecological Modelling, v. 157, p. 227–247.

Miranda, M.I., 1999, Clustering methods and algorithms, accessed January 2010 at *http://www.cse.iitb.ac.in/dbms/Data/Courses/CS632/1999/clustering/dbms.html*.

Mitsch, W.J., and Gosselink, J.G., 2007, Wetlands (4th ed.): New York, Wiley, 600 p.

Monte, J.A., 1978, The impact of petroleum dredging on Louisiana's coastal landscape: A plant biogeographical analysis and resource assessment of spoil bank habitats in the Bayou Lafourche delta: Baton Rouge, Louisiana State University, Ph.D. dissertation, 351 p., 57 figs.

National Park Service, 2009a, Barataria Preserve: Accessed October 3, 2009, at: *http://www.nps.gov/jela/barataria-preserve.htm*.

National Park Service, 2009b, Canal reclamation at Barataria Preserve: Environmental assessment Jean Lafitte National Historical Park and Preserve: National Park Service Technical Report, New Orleans, La.

Nayegandhi, Amar, Brock, J.C., and Wright, C.W., 2009, Small-footprint, wave-resolving lidar estimation of submerged and sub-canopy topography in coastal environments: International Journal of Remote Sensing, v. 30, no. 4, p. 861–878.

Nayegandhi, Amar, Brock, J.C., Wright, C.W., and O'Connell, M.J., 2006, Evaluating a small footprint, waveform-resolving LiDAR over coastal vegetation communities: Photogrammetric Engineering and Remote Sensing, v. 72, no. 12, p. 1407–1417.

Rousseeuw, P.J., and Leroy, A.M., 1987, Robust regression and outlier detection: New York, Wiley, 329 p.

Salovaara, K.J., Thessler, Sirpa, Malik, R.N., and Tuomisto, Hanna, 2005, Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery: Remote Sensing of Environment, v. 97, p. 39–51.

Sesnie, S.E., Gessler, P.E., Finegan, Bryan, and Thessler, Sirpa, 2008, Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments: Remote Sensing of Environment, v. 112, p. 2145–2159.

Shewchuk, J.R., 1996, Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator, *in* Lin, M.C., and Manocha, D. eds., Applied computational geometry. Towards geometric engineering: Berlin, Springer, p. 203–222.

Struyf, Anja, Hubert, Mia., and Rousseeuw, P.J., 1997, Integrating robust clustering techniques in S-PLUS: Computational Statistics and Data Analysis, v. 26, p. 17–37.

Urbatsch, L.E., Ferguson, D.M., and Gunn-Zumo, S.M., 2007, Vascular plant inventories of Jean Lafitte National Historical Park and Preserve, Barataria Preserve and Chalmette Battlefield: Baton Rouge, Louisiana State University, Department of Biological Sciences Technical Report, 109 p.

U.S. Geological Survey, National Wetlands Research Center, 2008, 2007 Louisiana coastal marsh–Vegetative type dataset: U.S. Geological Survey, accessed November 2010 at *http://sabdata.cr.usgs.gov/sabnet_pub/pub_sab_app.aspx?prodid=22780*.

Wagner, Wolfgang, Ullrich, Andreas, Melzer, Thomas, Briese, Christian, and Kraus, Karl, 2004, From single-pulse to full-waveform airborne laser scanners: Potential and practical challenges: International Archives on Photogrammetry and Remote Sensing, v. 35B, p. 201–206.