



## Georgia Statewide Assessment of 2009 Canopy



Huidae Cho<sup>1</sup>, Ph.D., Owen Smith, Jennifer McCollum

Institute for Environmental and Spatial Analysis, University of North Georgia

May 27, 2020

### Disclaimer

The work upon which this publication is based was funded in whole or in part through an Urban and Community Forestry grant awarded by the Southern Region, State and Private Forestry, U.S. Forest Service and administered by the Georgia Forestry Commission.

In accordance with Federal law and U.S. Department of Agriculture (USDA) policy, this institution is prohibited from discriminating on the basis of race, color, national origin, sex, age, or disability. (Not all prohibited bases apply to all programs.)

To file a complaint of discrimination, write USDA Director, Office of Civil Rights, Room 326-W, Whitten Building, 1400 Independence Avenue, SW, Washington, DC 20250-9410 or call 202-720-5964 (voice and TDD). USDA is an equal opportunity provider and employer.

### Abstract

Georgia is home to a diverse ecological system and it is important to monitor tree canopy changes to maintain such high biodiversity in the state. This project aims to classify canopy and non-canopy across the entire state of Georgia using the 2009 1-meter National Agricultural Imagery Program (NAIP) imagery. There are total 3,913 quarter quad NAIP tiles, so it is error-prone to process thousands of tiles manually. In this project, we proposed an efficient and automated workflow and developed a Python module called CanoPy. For canopy classification, we used the previously trained Feature Analyst models from the 2016 study of 2015 statewide canopy. Issues were found with water body detection, but they were believed to be due to the reuse of 2015 training data and would only account for less than 1% of the state's area according to our discussion with the GFC. An unweighted average accuracy of 87.6 % and an area-weighted average accuracy of 88.8% were attained for the entire state, which are both greater than the required 85% accuracy. The statewide percentage of 2009 canopy was 71%.

## Table of Contents

1. Introduction.....2

<sup>1</sup> Principal investigator. Email address: [hcho@ung.edu](mailto:hcho@ung.edu)

2. Materials and Methods .....	3
2.1. Physiographic Districts of Georgia .....	3
2.2. Improved Workflow Using ArcPy .....	3
2.2.1. Reprojection of the Individual NAIP Tiles .....	4
2.2.2. Feature Extraction .....	6
2.2.3. Seamless Patching of the Output Tiles .....	7
2.3. Ground Truthing .....	9
3. Results and Discussion .....	10
3.1. Ground Truthing Results .....	10
3.3. Classification Issues .....	12
3.4. Future Considerations .....	12
4. Conclusions .....	13
Acknowledgments .....	14
References .....	14

## 1. Introduction

Georgia is home to a massively diverse ecological system and, when the overall number of species within Georgia is compared with that in the other states, Georgia ranks sixth in biodiversity (GA EPD, 2009). Additionally, the different ecological zones all have different types of diverse forests with hardwood forests in northern Georgia, mixed forests in the Piedmont district, and the Coastal plain consisting of longleaf pine forests (GA EPD, 2009). It is important to maintain and monitor the environment that allows for such high levels of biodiversity to exist. To aid in monitoring tree canopy, data must be created periodically, but the creation of a large-scale canopy dataset can be a time-consuming and difficult task. To address this computational inefficiency in generating such canopy dataset, Textron Systems’ Feature Analyst and the Python programming language were leveraged for Artificial Neural Network classification and process automation, respectively. The Python language allowed for the creation of a new method aimed at mitigating potential errors in the data and will allow for improved reproducibility during future iterations.

The Institute for Environmental Spatial Analysis (IESA) at the University of North Georgia (UNG), funded by the Georgia Forestry Commission (GFC), developed a method with which to classify canopy at a statewide scale utilizing Feature Analyst which is tightly integrated with ArcMap 10.5. The National Agricultural Imagery Program (NAIP) imagery (USDA, 2020) used to classify canopy is made available approximately every four years. While it cannot match the temporal frequency of satellites such as Landsat and MODIS, the NAIP imagery is taken during seasons in which agriculture is growing in the United States ensuring similar characteristics between datasets (USDA, 2020). Additionally, factors such as the quality control that NAIP undergoes and a high resolution (1m) ensures a result with high accuracy. Using the method developed with the NAIP imagery, a baseline canopy dataset for Georgia in 2009 was created for

future comparisons with other canopy datasets created with the same method. The creation of the dataset was started as a follow up to the 2016 study commissioned by the GFC to create a canopy dataset for the year 2015. The tools and methods for the creation of the 2009 dataset were pulled directly from the 2016 study because, in the second phase of the project, the results of both datasets were to be compared to study canopy change across the state. The machine learning models used for both studies were the same as well. The level of parallelism between the two studies was to ensure that the 2009 dataset would be a like-for-like comparison in terms of the way it was created so as to allow for better change detection. However, while the models of the 2016 study were found to be highly accurate (Bailey and Bailey, 2019), the previous method was improved upon by our team as detailed below.

This document details in depth the process undertaken in the completion of the first phase of the GFC canopy assessment project. The goal of this phase was to create a baseline canopy dataset of the state of Georgia for the year 2009 using the 1-meter NAIP imagery. The results will be used to monitor canopy change across years in next phases.

## 2. Materials and Methods

### 2.1. Physiographic Districts of Georgia

The state of Georgia consists of 24 physiographic districts as shown in Figure 1. These districts were used to split the study area into 24 separate processing regions, each utilizing a model trained separately from the others. This separation of the state into multiple regions would allow for better results and shorter computation times than if the entire state had been utilized at once or even if the five physiographic provinces of Georgia had been used. The 2009 NAIP imagery is separated into 3,913 quarter quad (QQ) tiles with a size of 3.75 by 3.75 minute per tile plus an extra 300-meter buffer on each side of the QQ. Dissemination of the data is subsequently separated by the 24 districts.

### 2.2. Improved Workflow Using ArcPy

The new method developed for this project looked to improve the following aspects: inconsistent tile shifting, boundary artifact creation, and unknown mosaic order leading to inconsistent results within the 300m buffer area of the NAIP imagery. The new method created is broken into three main steps: (1) reprojection of the individual NAIP tiles, (2) feature extraction, and (3) seamless patching of the output tiles. These steps would be a time-consuming and tedious task with an increased chance of error if left to manual inputs without the use of any process automation. To achieve the level of automation and consistency desired for our study, a Python module was created titled CanoPy. It is available freely at <https://gislab.isnew.info/canopy/start>. The Python programming language was chosen because of the availability of ArcMap's Python Application Programming Interface (API) allowing for easy integration with the software already in use. CanoPy consists of separate functions created for each step of the data creation process with the exception being feature extraction because Textron Systems' proprietary software has no readily available Python API that fits the needs of the system. In addition to the data pre-processing functions, CanoPy contains several post-processing analysis functions such as ground truth point creation and GeoTIFF to shapefile conversion. CanoPy is split into two main files including the function file canopy.py containing all created functions, and the configuration file canopy\_config.py. The configuration

file contains inputs for all of the data paths, such as the name of the NAIP QQ shapefile and the physiographic district shapefile, needed to run the CanoPy functions.

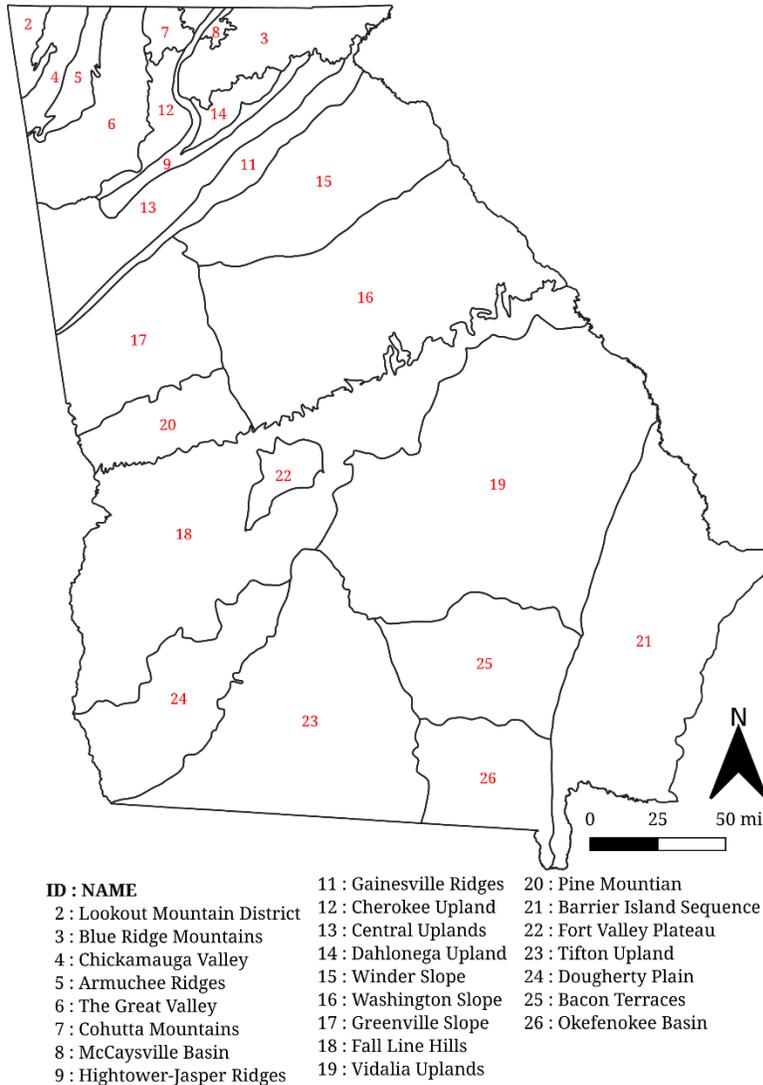


Figure 1. Physiographic districts of Georgia.

### 2.2.1. Reprojection of the Individual NAIP Tiles

The projection used for the dataset is the USA Contiguous Albers Equal Area Conic USGS version projection (WKID 102039) which follows the GFC's convention. Additionally, as the NAIP tiles are in Universal Transverse Mercator (UTM) projections initially and Georgia is split into two UTM zones, this reprojection step will ensure that the final dataset is in the same projection across the state. The first tile reprojected will have no reference snapping applied, but every reprojected tile after will use this first raster as its reference. This snapping will remove any inconsistent tile shifting which would lead to inconsistencies when comparing or mosaicking.

### Assigning Physiographic District IDs to NAIP QQ Tiles

As processing regions are split into the 24 physiographic districts of Georgia, the first step taken was to formulate a way that would allow for only the NAIP imagery in a specified district to be processed at a time. This problem was approached by utilizing spatial relation between the Georgia physiographic district shapefile which contains numerical identifiers (IDs) for each district, and the NAIP QQ seamline shapefile which contains key identifying information for each NAIP QQ raster file. Firstly, a new field is added to the NAIP QQ shapefile titled “PHYREGS.” Each physiographic district is then iterated over and, by using ArcPy’s Search Cursor and “Select Layer By Location” function, all NAIP QQs within each physiographic district are selected and the PHYREGS field is populated with the numerical ID of the physiographic district. Some NAIP QQs can have multiple physiographic IDs when the boundaries of multiple physiographic districts split the QQs. To account for multiple physiographic districts, the formatting of the PHYREGS field is not numerical but is text separated by commas (e.g., [3,15,...]) enabling NAIP QQs to be used for multiple districts.

By assigning the physiographic district IDs in the manner described above, each subsequent function can leverage the physiographic district IDs as inputs allowing for efficient processing on a district-by-district basis. In addition, this method also allows for a clean and tidy output folder structure to be created by separating out each district's outputs automatically and ensuring no duplicates are created as shown in Figure 2.

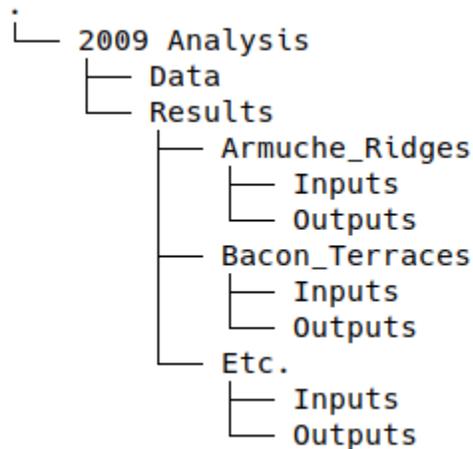


Figure 2. Folder structure of 2009 analysis.

### Reprojecting Input Tiles

To ensure all final deliverables are in the same projection and within the same raster grid, each individual tile to be classified is reprojected and snapped to the same raster before undergoing classification. First, all NAIP QQs within the desired district are selected using the PHYREGS attribute assigned to the NAIP QQ tile. The QQ feature selection ensures that, when ArcPy’s Search Cursor iterates over the NAIP QQ shapefile, it only iterates over the QQs that have been selected. The filename of each tile is read from the NAIP QQ shapefile and parsed together with the NAIP directory path specified in the Canopy configuration file. Each tile is then reprojected to the USA Contiguous Albers Equal Area Conic USGS version projection as specified by the GFC. However, each reprojected tile does not take the place of the original NAIP

imagery tile. Instead, two folders are created, a folder titled after the physiographic district and a subfolder titled “Inputs” under the district folder. The reprojected NAIP tile is saved within the input folder of its corresponding district with the prefix of “r” to denote Reprojected. Prefixes are added to every piece of data created throughout the process allowing for the end-user to be able to determine at which step the data was created. In this case, the prefix of “r” denotes that the output data was created by the reprojection function. Additionally, the prefix added to each tile allows other functions to read the proper NAIP imagery. For the snap raster, each function uses the arcpy.env settings to declare the snap raster.

### 2.2.2. Feature Extraction

Feature extraction or canopy classification was done using Feature Analyst. Feature Analyst is an Object-Based Image Analysis (OBIA) software package for Esri ArcMap. It is used to extract features from high-resolution images for geospatial analysis (Textron Systems, 2020). This ArcMap extension was used during the previous 2016 study with support from Textron Systems. For that reason, Feature Analyst was also used in this study in order to maintain consistency. Feature Analyst models are stored in AFE (Automated Feature Extraction) model files. The AFE file is a proprietary file format unique to Feature Analyst and is a wrapper that is created and exported out after a model is generated. This file format contains all steps and training data required for the model to run. The cohesive nature of the file format allows for models to be easily reused as was done for this study with the use of the previous study’s models (Textron Systems, 2020). Once the Feature Analyst software is run, it creates a file for both the source and output layers. The output layer file uses an AFE extension and stores all the settings. Therefore, if the extraction process produces good results, the AFE file can be reused on other images (Textron Systems, 2020). If it does not produce good results, the model can be refined as many times as necessary to achieve best results. See Figure 3 for the general workflow of Feature Analyst. Figure 4 shows the structure of an AFE file. As the previous study created a model for each physiographic district, there are 24 distinct models.



Figure 3. General workflow of Feature Analyst.

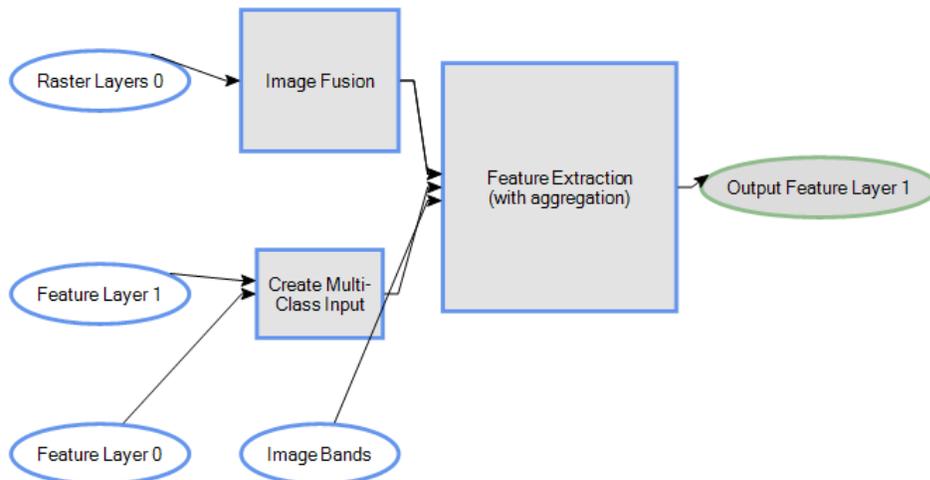


Figure 4. Structure of an AFE file.

Following the reprojection of the input NAIP tiles, classification is now ready to occur. For classification, the batch classification tool in Feature Analyst was used where the AFE file is loaded in along with all reprojected files in the district input folder. An “Outputs” folder needs to be created by the user within the district result folder. The classified tiles will be saved in the Outputs folder and all subsequent CanoPy functions will be read and output data into the same folder.

### 2.2.3. Seamless Patching of the Output Tiles

#### *Overlapping Tiles and Linear Artifacts*

The unknown mosaic order from the 2016 study initially created issues with consistency within the 300m buffer area. ArcGIS’s “Mosaic To New Raster” tool uses the LAST method by default, which overwrites any cells from previous tiles with those from the last tile. However, there was no way to know which order the tiles were overlapped. An algorithm was developed that would help determine the mosaic order, but the percentage of error created by both the linear artifacts and the overwriting of cells by the LAST mosaic method in the 2016 study led us to approach mosaicking differently. Instead of attempting to recreate the mosaic order of the 2016 study, we would instead use the NAIP QQ seamline shapefile provided by the USDA to remove both the linear artifacts and the need for a mosaic order in one step. After the individual tiles are processed by Feature Analyst, the linear artifacts created along the boundaries must be removed. Using the filename attribute contained within each NAIP QQ polygon, the corresponding tile is then clipped by the NAIP QQ polygon. The clipped NAIP tile removes the 300m buffer of each classified NAIP tile and removes the linear artifacts created by Feature Analyst on the edges of the buffer. The removal of the 300m buffer also removes any overlap between tiles, subsequently removing any cell overwriting by the mosaicking method. Figure 5 shows differences in results between the new and previous methods.



Figure 5. Differences in results between the new (left) and previous (right) methods.

### Conversion of AFE Outputs to Final Tiles

Not all of the AFE models output the data in the same format. Discrepancies include the creation of shapefiles as opposed to GeoTIFF files and GeoTIFF files with the values of 1 (non-canopy) and 0 (canopy), which will be handled later, as opposed to 0 (canopy) and 1 (canopy). To change the outputs to the required output of a 2-bit GeoTIFF file with the values 0 (non-canopy) and 1 (canopy), each file after classification is iterated over using ArcPy's SearchCursor. As CanoPy passes over each tile within the processing district, it converts all shapefiles in the district output folder to GeoTIFF files with ArcPy's "Feature To Raster Conversion" function and uses the shapefiles CLASS\_ID field as the field to populate each cell. Additionally, as CanoPy iterates over each file, it reclassifies each GeoTIFF file to the required values of 0 and 1. The output files are saved with the prefix "fr" denoting that the tiles are Final Reprojected tiles.

### Clipping Final Tiles

Following the batch classification and conversion, the data must undergo additional processing to get rid of edge artifacts and remove the necessity of determining the mosaicking order of the 2016 study. To automate the process, each tile with the prefix of "fr" is selected by iterating over each NAIP QQ polygon within the processing district and reading the filename for each tile with the prefix attached. Each "fr" tile is clipped to its respective QQ polygon with ArcPy's "Extract By Mask" function, effectively removing any overlap between adjacent tiles and removing the edge artifacts of each tile. Outputs are saved with the prefix of "cfr" denoting Clipped Final Reprojected tiles.

### Mosaicking Clipped Final Tiles

Discrepancies within the output formats of each AFE model and the subsequent conversion between formats and values lead to 10 districts in which the values were inverted to 0 for canopy and 1 for non-canopy. The inverted districts were initially noticed after the final physiographic district mosaics. To fix the inversion, the value of the inverted GeoTIFF file is subtracted from 1. That is, 0 (canopy) becomes  $1-0 = 1$  (canopy), and 1 (non-canopy) becomes  $1-1 = 0$  (non-canopy). The correction method applied to the final district mosaic was suitable for the final deliverables of this study. However, future analysis will apply the correction of inverted districts earlier in the process to each individual tile.

The final step of the method is to mosaic the “cfr” tiles and clip the mosaicked tiles to the physiographic district boundaries. This step iterates over all files in the specified district output folder using the Search Cursor to read the required filenames from the NAIP QQ shapefile. The required rasters are mosaicked and saved as a 2-bit GeoTIFF file. The 2-bit file size allows for the file size of the mosaicked and clipped outputs to be significantly smaller than if it was saved in its original 32-bit file size. This conversion is possible because the final rasters have binary values including 0 and 1 for non-canopy and canopy, respectively. After the 2-bit mosaicked raster is created, it is clipped to the specified physiographic district boundary.

### 2.3. Ground Truthing

The assessment of each district’s accuracy was undertaken at a visual scale of 1:2500. The number of ground-truthing points for each district was decided by linear interpolation across the districts based on the area with a maximum of 400 points and a minimum of 200 points. To randomly generate the points for each district, ArcPy’s “Create Random Points” function was used to create a temporary point shapefile within the boundaries of each specified physiographic district. A field in the temporary point layer was created titled “GT” which will be populated by the values of the classified raster. To populate the GT field, already available ArcPy functions such as “Extract By Point” could not be utilized as it would only pull the value from one point and would require large amounts of manual input to the function. It was decided to approach the population of the GT field by using NumPy, a Python library for working with large scale matrices (e.g., raster datasets). While there are several options for reading raster datasets as NumPy arrays, ArcPy has a function for reading raster datasets as NumPy arrays. Initially, the ground-truthing method was to read the values out of the final mosaicked and clipped raster. However, issues arose out of the computational memory needed to store each district’s array and even the smallest district was too large for ArcPy to handle. To work around the computational limitations, it was decided to use the classified and clipped individual tiles to read the data of each point. To be able to get the required filename of each tile, a spatial join is carried out to join the data of the temporary shapefile and the NAIP QQ shapefile. The temporary point shapefile is then deleted. Reading the raster values at each point in a NumPy array is accomplished by converting the coordinates of each point on the tile to row and column. By using the X and Y values of each point, the value for each point on the classified NAIP QQ tile was found and the GT field is populated. All fields other than the OID, Shape, and GT fields were subsequently removed to make the final shapefile as tidy and small as possible. For the visual inspection, a short integer field called “VISUAL” was added manually to each point shapefile when it was time to inspect that district.

Additional fields were created in the ground-truthing shapefile titled “CHECK” and “ERROR.” The VISUAL field was then filled manually. To cut down on the time it would take to do a visual inspection, a Python function was created to automatically load the NAIP imagery that contains a point, into ArcMap or ArcGIS Pro. The use of this function cut down on time it would take to find each required NAIP tile out of the 3,913 available tiles. After the required imagery was loaded, the visual inspection viewed each point at a scale of 1:2500 and entered either a 0 for non-canopy or 1 for canopy into the VISUAL field for the point. After every point within a district is checked, the CHECK field is populated with three possible values. It can either be 0 meaning that both the GT and VISUAL values are the same, 1 (a false positive) meaning that the VISUAL was canopy, but the GT was non-canopy, or 2 (a false negative) meaning that the VISUAL

was non-canopy, but the GT was canopy. The error was calculated from the CHECK field for each point by inserting a value 1 when a point's CHECK value is greater than 0.

### 3. Results and Discussion

#### 3.1. Ground-Truthing Results

Ground-truthing inspection yielded positive results. The maximum district accuracy was 99.0% and the minimum was 76.6% with only three districts having accuracies below 80.0%. Two averages were calculated including one taken as the unweighted average and one weighted by the district area. Both averages showed a high accuracy percentage with the unweighted average equaling 87.6% accuracy and the area-weighted average equaling 88.8% accuracy. Both accuracies are greater than the required 85.0%. Individual accuracies and data can be seen in Table 1.

Table 1. All accuracies for individual districts and aggregated accuracies. The average and standard deviation of the accuracy is 87.6% and 5.9%, respectively. The area-weighted average accuracy is 88.8%.

Physiographic District	Area (mi <sup>2</sup> )	Total Points	Error Points	% Accuracy	Area-Weighted % Accuracy
Armuchee Ridges	477	208	37	82.2	0.7
Bacon Terraces	2,644	256	30	88.3	3.9
Barrier Island Sequence	6,009	331	43	87.0	8.8
Blue Ridge Mountains	1,444	229	4	98.3	2.4
Central Uplands	2,141	245	33	86.5	3.1
Cherokee Upland	547	209	9	95.7	0.9
Chickamauga Valley	536	209	49	76.6	0.7
Cohutta Mountains	267	203	2	99.0	0.4
Dahlongega Upland	364	205	32	84.4	0.5
Dougherty Plain	2,560	254	34	86.6	3.7
Fall Line Hills	7,732	370	36	90.3	11.7
Fort Valley Plateau	475	208	22	89.4	0.7
Gainesville Ridges	991	219	39	82.2	1.4
Greenville Slope	2,544	254	34	86.6	3.7
Hightower-Jasper Ridges	592	210	24	88.6	0.9
Lookout Mountain District	334	205	27	86.8	0.5
McCaysville Basin	149	200	22	89.0	0.2
Okefenokee Basin	1,923	240	51	78.8	2.5
Pine Mountain	1,319	227	31	86.3	1.9
The Great Valley	1,795	237	53	77.6	2.3
Tifton Upland	5,761	326	13	96.0	9.3
Vidalia Uplands	9,104	400	20	95.0	14.5
Washington Slope	6,386	340	56	83.5	9.0
Winder Slope	3,390	273	34	87.5	5.0
Total	59,485	6,058	735	87.6%	88.8%

### 3.2. Statewide Canopy in 2009

Figure 6 shows the canopy analysis results for individual physiographic districts for the year 2009. Overall, the percentage of statewide canopy was 71%. The percentage of canopy for individual districts ranged between 47% for Dougherty Plain and 99% for Cohutta Mountains. Canopy area percentages for individual districts can be seen in Table 2.

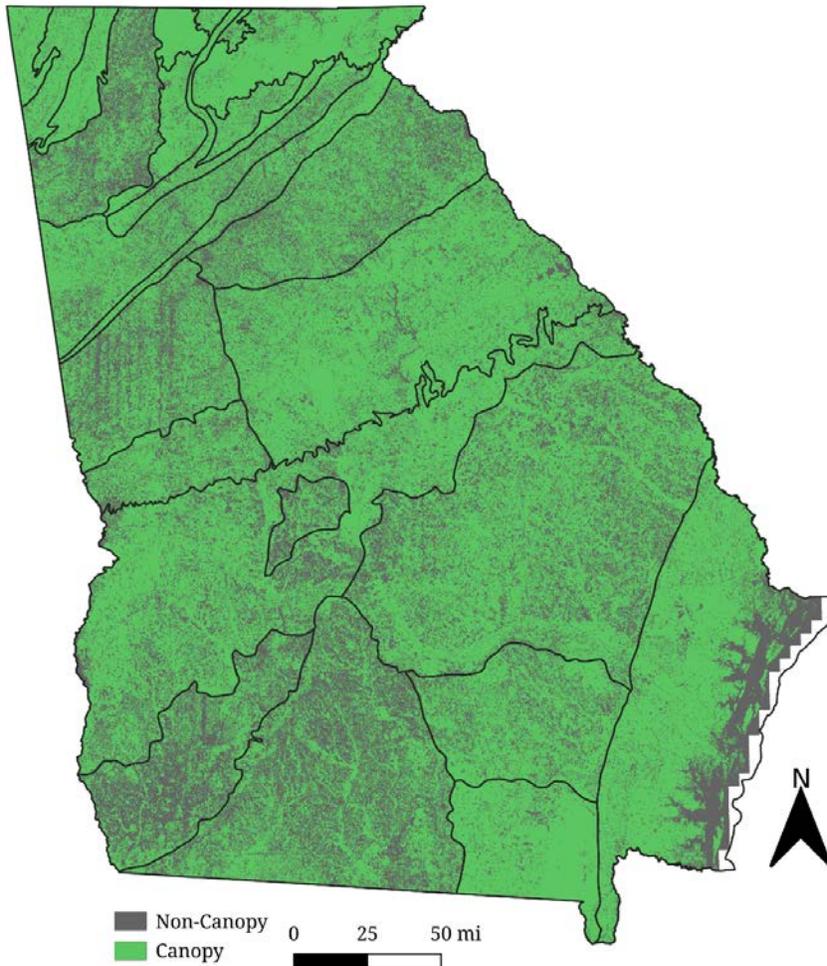


Figure 6. 2009 statewide canopy dataset for Georgia. Black polygons indicate physiographic district boundaries.

Table 2. Area percentages that are canopy for individual districts. The total percentage of area that is canopy is 71% of the state.

Physiographic District	Area (mi <sup>2</sup> )	Canopy (mi <sup>2</sup> )	% Canopy
Armuchee Ridges	477	445	93
Bacon Terraces	2,644	2,000	76
Barrier Island Sequence	6,009	4,077	68
Blue Ridge Mountains	1,444	1,333	92
Central Uplands	2,141	1,602	75

Cherokee Upland	547	472	86
Chickamauga Valley	536	477	89
Cohutta Mountains	267	263	99
Dahlonega Upland	364	296	81
Dougherty Plain	2,560	1,201	47
Fall Line Hills	7,732	5,839	76
Fort Valley Plateau	475	264	56
Gainesville Ridges	991	757	76
Greenville Slope	2,544	1,579	62
Hightower-Jasper Ridges	592	505	85
Lookout Mountain District	334	310	93
McCaysville Basin	149	127	85
Okefenokee Basin	1,923	1,765	92
Pine Mountain	1,319	938	71
The Great Valley	1,795	1,004	56
Tifton Upland	5,761	3,157	55
Vidalia Uplands	9,104	6,531	72
Washington Slope	6,386	5,316	83
Winder Slope	3,390	2,185	64
Total	59,484	42,443	71%

### 3.3. Classification Issues

Throughout the inspection phase of the study, some issues were found. The first issue found was the previously discussed inversion of canopy values. The inversion can be easily fixed through CanoPy early in the process on a tile-by-tile basis and, for future iterations of the study, it will be. It is still, however, an issue that needs to be kept in mind and a quick manual check throughout the process can only further ensure the quality of the data.

The second issue discovered was a lack of accurate classification for bodies of water as shown in Figure 7. This misclassification of water bodies is believed to be due to the reuse of datasets created for 2015 and not the 2009 data which this study was conducted on. Additional processing suggested to rectify the issue was the use of the National Hydrography Dataset (NHD) for water truthing, or rather using the NHD to compare areas where water should be in certainty and, if it is classified as canopy, the value is converted to non-canopy. However, the water truthing was not undertaken after a discussion with the GFC, as it is believed that the affected area would be less than 1% of the total area of Georgia.

### 3.4. Future Considerations

The development of CanoPy is still ongoing with the goals of improving accessibility and the outputs. As the NAIP imagery is transitioning to a higher sub-meter resolution imagery (i.e., 0.6m), computer resource management will be even more important for future iterations. Future iterations could produce datasets potentially about 3 times larger than previous 1m datasets (i.e.,  $[1\text{m}/0.6\text{m}]^2=2.8$ ). Because of the increased data size, it will become important to downsize data to 2-bit wherever possible. To achieve this efficiency in data storage, CanoPy will need to be further developed to be able to handle datasets in different

resolutions. Additionally, comparison algorithms will need to be proposed to account for the resolution differences as naïve cell-by-cell comparisons will not be adequate. A moving window algorithm might be more ideal for future comparisons. A comparison coefficient that can detect spatial patterns may be better for future change analyses in different resolutions.

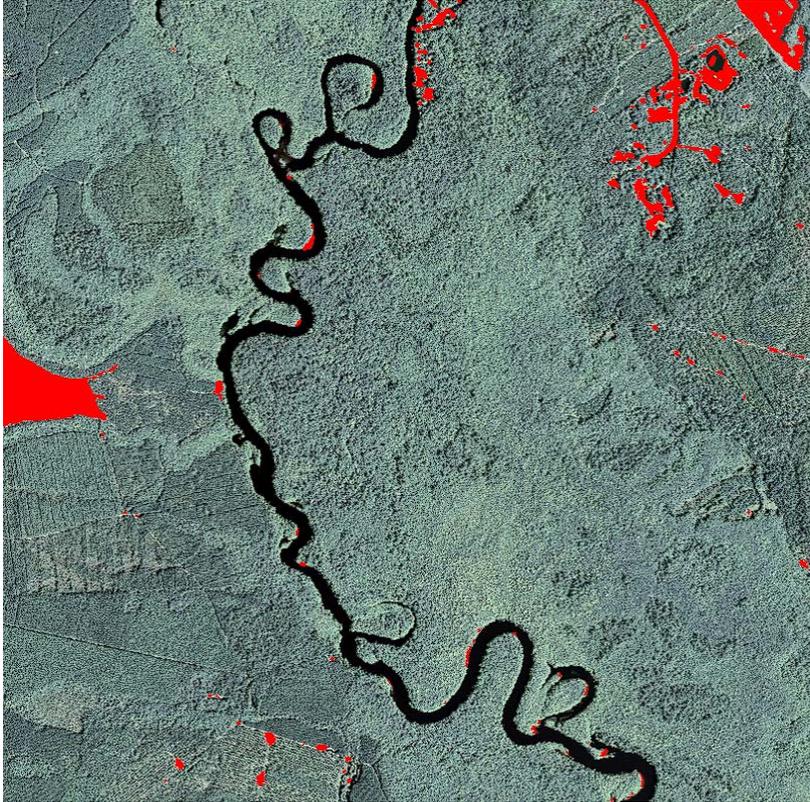


Figure 7. Misclassification of water bodies. Red shows areas that were classified as non-canopy. The river flowing centrally is not classified as non-canopy.

#### 4. Conclusions

Canopy classification is an important but time-consuming process at a statewide scale. CanoPy, a Python module, was developed to automate and streamline an improved method for canopy classification. The automated process involves three main steps including (1) reprojection of the individual NAIP tiles, (2) feature extraction, and (3) seamless patching of the output tiles. The process effectively creates data free of any errors that would be created because of anything other than the classification. Errors that could arise, such as inverted values, can be rectified with CanoPy. Although issues with water body detection were discovered that are believed to be due to the reuse of 2015 training data, it would only account for less than 1% of the state's area according to our discussion with the GFC. Because of the minimal area with the water misclassification issue, water truthing was not undertaken to rectify the issue. Visual inspection of the data was performed and found an unweighted average accuracy of 87.6 % and an area-weighted average accuracy of 88.8% for the entire state, which are both greater than the required 85% accuracy. The physiographic districts with the minimum and maximum canopy percentages were Dougherty Plain (47%) and Cohutta Mountains (99%), respectively. The overall statewide percentage of canopy for the year 2009 was 71%.

## Acknowledgments

This project was funded in whole through an Urban and Community Forestry grant awarded by the Southern Region, State and Private Forestry, U.S. Forest Service and administered by the Georgia Forestry Commission. The authors would like to thank Dr. Allison Bailey and her students for having trained the canopy classification models for the entire Georgia and providing us with data from their 2015 analysis.

## References

Bailey A. J., Bailey C. O., Jr., 2019. Using Feature Analyst & NAIP Imagery to Conduct a Statewide Tree Canopy Assessment of Georgia. *Forestry Research Engineering: International Journal* 3 (1), 15–18.

Georgia Environmental Protection Division (GA EPD), 2009. State of Georgia's Environment 2009. [https://epd.georgia.gov/sites/epd.georgia.gov/files/related\\_files/site\\_page/Objective%202.pdf](https://epd.georgia.gov/sites/epd.georgia.gov/files/related_files/site_page/Objective%202.pdf). Accessed on March 21, 2020.

Textron Systems, 2020. Feature Analyst Transform Imagery into Geospatial Understanding. Textron Systems. <https://www.textronsystems.com/products/feature-analyst>. Accessed on April 18, 2020.

United States Department of Agricultura (USDA), 2020. NAIP Imagery. <https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/>. Accessed May 25, 2020.