Urban Forestry Program

Accuracy Assessment of 2010 Tree Canopy Data

Prepared by Alan Halter

2014

AUSTIN PARKS & RECREATION
Cultural Places, Natural Spaces

Commission for Accreditation of Park and Recreation Agencies
CAPRA ACCREDITED

URBAN FORESTRY
CITY OF AUSTIN

CITY OF AUSTIN
FOUNDED 1839

OVERVIEW

- The draft 2010 tree canopy feature class is highly accurate at representing tree canopy on the ground
- Out of a total 200 sample points, only 14 errors were found
- 93% of pixels classified as tree canopy in the GIS feature class were actual tree canopy in the original NAIP aerial image
- 88% of all tree canopy in the original NAIP aerial image was classified as tree canopy in the GIS feature class
- The draft 2010 tree canopy feature class shows a strong agreement and statistically significant geographic representation of tree canopy in Austin (Kappa value = 0.849)

INTRODUCTION

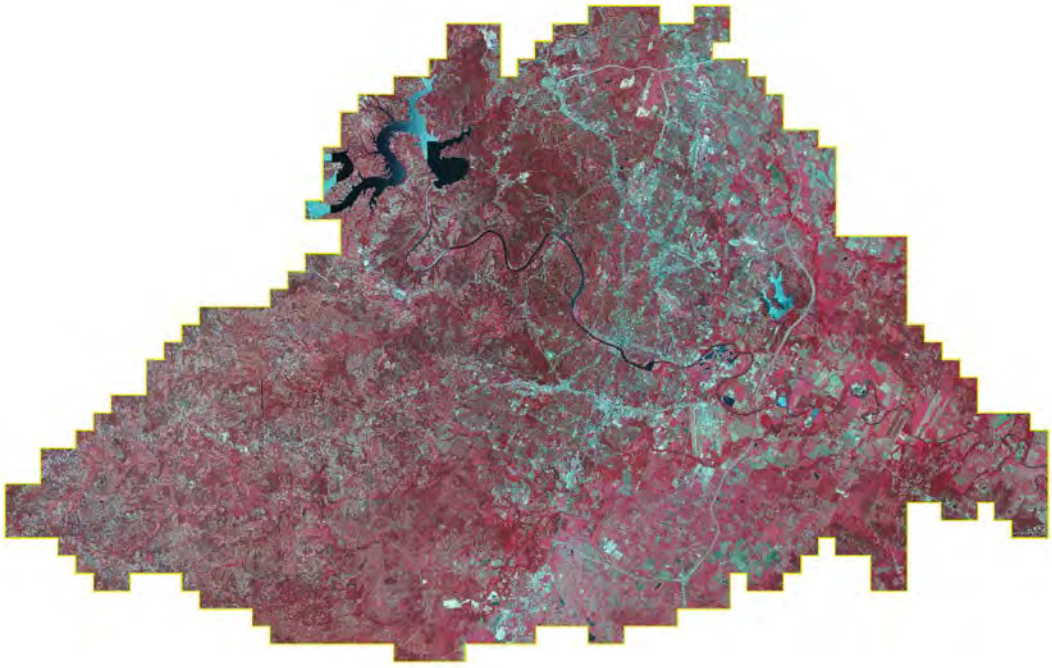
This report provides methods and findings of an accuracy assessment conducted on the draft 2010 tree canopy data created by the City of Austin's Watershed Protection Department (WPD).

The draft 2010 tree canopy data is a Geographic Information Systems (GIS) feature class showing the spatial distribution of tree canopy across the entire City of Austin jurisdiction. Tree canopy refers to "the layer of leaves, branches, and stems of trees that cover the ground when viewed from above" (USDA, 2009).

In this report, accuracy means correctness in depicting what is truly found on the ground. An accuracy assessment is a field validation technique that ascertains the quality of information derived from remotely sensed imagery (Congalton & Green, 1999). Map data or classified data (i.e. the draft 2010 tree canopy data) is evaluated against groundtruth data or reference imagery (i.e. 2010 NAIP aerial imagery) wherein the reference imagery is assumed to be 100% accurate. Results from this assessment determine the accuracy of the classified map data by quantifying the level of error inherent in its creation. These results not only help users understand how well mapped tree canopy estimates actual tree canopy on the ground, it also helps to show where improvements to the data extraction process may be made.

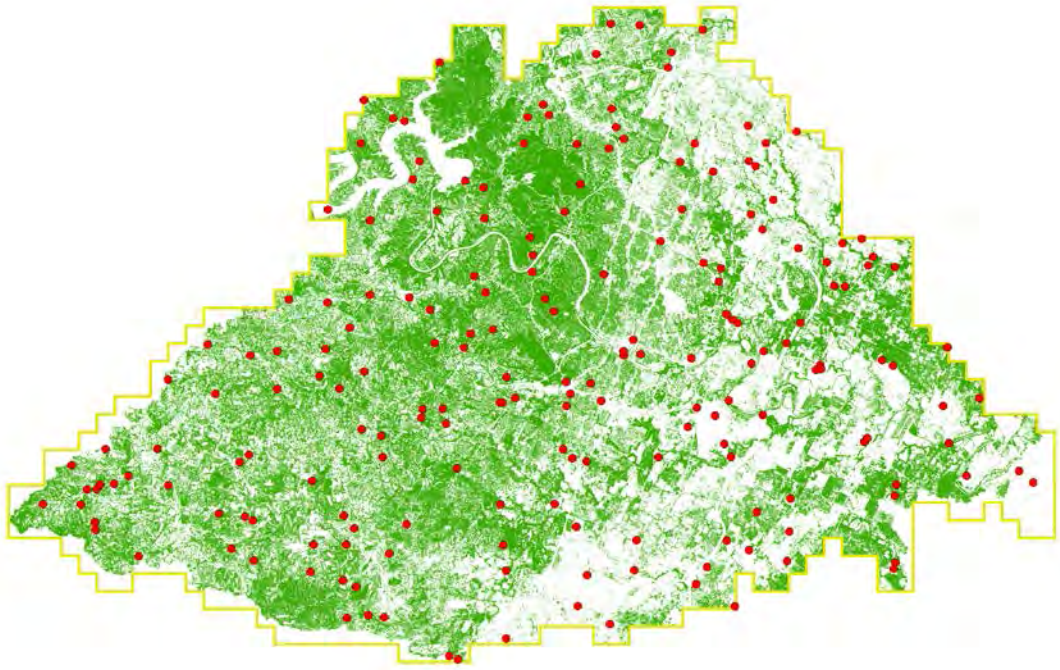
The tree canopy data was extracted from aerial imagery provided by the USDA's National Agriculture Imagery Program (NAIP). See Figures 1 and 2. The USDA acquires aerial imagery during the "leaf-on" growing season to monitor agricultural production, among other purposes. The imagery is taken at a 1-meter resolution and is orthorectified (spatially corrected) to produce an accurate representation of the Earth's surface. Spatial accuracy is ensured through referencing both existing imagery and ground control points in which "all points fall within 6 meters of true ground at a 95% confidence level" (USDA, 2013).

Figure 1 | NAIP Aerial Imagery (2010) for Austin, Texas



Note: False color image showing healthy vegetation (red) and urbanized areas (gray). Tree canopy was extracted from this image to create the draft 2010 tree canopy feature class below.

Figure 2 | Draft 2010 Tree Canopy Feature Class (2010) for Austin, Texas



Note: Draft 2010 tree canopy feature class (green) and the 200 sample points (red).

METHODS

Sample Design

A stratified random sample was chosen for which each map class (e.g. areas covered in tree canopy vs. areas not covered in tree canopy) was sampled independently and randomly with sample points determined by the proportion of land area covered by each map class. Out of a 739,213 acre study area, tree canopy in the GIS feature class covers 35.5% of the study area (262,911 acres) while area absent of tree canopy covers 64.5% of the study area (476,302 acres).

Sample Size Determination

As a general rule of thumb, Congalton & Green (1999) and Lillesand et al. (2004) suggest a minimum sample size of n=50, whereas similar studies show a range of sample sizes from n=50 to n=1,000. A total sample size of n=200 was decided for this assessment based on 1) a sample determination equation, 2) a review of similar studies, and 3) time and resources available to complete the assessment.

The number of sample points was calculated using the following equation based on binomial probability theory (Skirvin, et al., 2004):

$$n = \frac{Z^2 pq}{E^2}$$

Where:
n = number of samples,
p = expected or calculated accuracy (%)
q = 100 – p (expected error)
E = allowable error
Z = standard normal deviate for the 95% two-tail confidence level = 1.96

Preliminary analysis of the 2010 canopy data showed an overall accuracy of 85%, thus the expected accuracy was set at 85% and allowable error was set at 5%. Therefore, q = 0.15. From this calculation the sample size equated to n=196. This number was rounded to n=200.

The 200 sample points were created using ArcMap's Create Random Points tool. Points were divided into each map class based on the area proportions stated above. Since tree canopy covers 35.5% of the study area, 35.5% of the tree canopy sample points (n=71) were selected. Since areas absent of tree canopy cover 64.5% of the study area, 64.5% of non-tree canopy sample points (n=129) were selected.

Error Matrix

The 200 sample points were analyzed amongst two GIS users at the City of Austin's Urban Forestry Program. Accuracy was determined for each point by heads-up visual interpretation. The user compared the draft 2010 tree canopy data (i.e. the classified data) to the original NAIP 2010 imagery (i.e. the reference data) at each point's location to assess agreement between what was derived from the original aerial imagery and what truly exists on the ground. All points were assessed and reviewed multiple times by both users for quality control.

RESULTS

The comparisons were summarized in an error matrix (Figure 3) to calculate accuracy results.

Figure 3 | Error Matrix

		Reference Data (NAIP Aerial image)		
		Canopy	Not Canopy	**Row Total**
Classified Data (2010 Tree Canopy Feature Class)	Canopy	**66**	5	**71**
	Not Canopy	9	**120**	**129**
	Column Total	**75**	**125**	**200**

Overall Accuracy	**93%**
Kappa	**0.849**

Note: Accuracy = (Number of samples classified correctly / Total number of samples)*100
Diagonals (gray) represent correctly classified samples
Off-diagonals represent misclassified samples

Producer's Accuracy

Canopy = 66/75 = 88%

Not Canopy = 120/125 = 96%

User's Accuracy

Canopy = 66/71 = 93%

Not Canopy = 120/129 = 93%

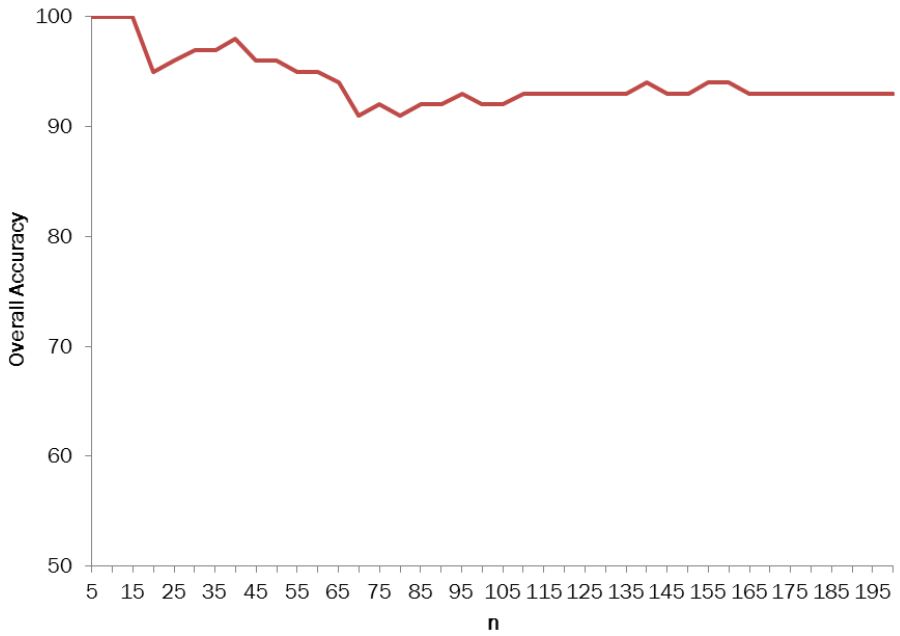
Overall Accuracy

The 200 sample points display an overall accuracy of 93% meaning that 93 out of 100 times, a point on the 2010 tree canopy map corresponds correctly with what was truly on the ground in 2010 (see Figure 3). This compares closely to overall accuracy findings of similar tree canopy data in various U.S. cities (Davey Resource Group, 2012; Davey Resource Group, 2011; Texas Trees Foundation, 2010).

- Tukwila, Washington (95.6%)
- Port Angeles, Washington (94%)
- Dallas, Texas (94.8%)

Figure 4 shows a running total of overall accuracy as the GIS users conducted their assessments. As the number of assessed sample points increased, the overall accuracy decreased to a point that eventually hovered around 93% beginning at n=70 and ending at n=200.

Figure 4 | Running Total of Overall Accuracy



User's Accuracy and Producer's Accuracy

Overall accuracy is an average and does not provide information regarding error between classes. User's accuracy displays errors of commission in a particular class (i.e. including an area in a category when it does not belong to that category), while producer's accuracy displays errors of omission in a particular class (i.e. excluding an area from the category to which it belongs). In other words, user's accuracy shows how many points on the map are actually what they say they are. For this assessment, the user's accuracy for non-canopied areas (93.02%) is slightly better than the canopy class (92.96%). On the other hand, the producer's accuracy for non-canopied areas (96%) is greater than the producer's accuracy of the canopy class (88%). Of importance is that, although 88% of the canopied areas have been correctly identified as canopy, 92.96% of the areas called canopy on the map are actually canopy on the ground. Of equal importance is that, although 96% of the non-canopied areas have been correctly identified as non-canopy, 93.02% of the areas called non-canopy on the map are actually non-canopy on the ground.

Kappa

The Kappa statistic shows the difference between actual agreement in the error matrix (i.e. the gray diagonals in Figure 3) and the agreement expected by chance (i.e. the row and column margin values in Figure 3) (Congalton & Green, 1999). A Kappa value greater than 0.80 represents a strong agreement, and can determine if values contained in the error matrix represent a statistically significant result better than a random result (Landis & Koch, 1977; Jensen, 1996). The Kappa value for this assessment was calculated at 0.849 meaning there is 84.9% better agreement than by chance alone; therefore the map data sufficiently represents what truly existed on the ground in Austin in 2010. Kappa findings of similar tree canopy data in other U.S. cities were 0.9315 in Tukwila, Washington and 0.85 in South Bend, Indiana (Davey Resource Group, 2012; Davey Resource Group, 2011).

Conclusion

The four main measures of accuracy each show high numeric values in their respective levels of accuracy. Overall accuracy is within 10 percentage points of perfect and within range of other U.S. cities. Both user's and producer's accuracies are greater than 85% showing a high likelihood that both the map producer and users will find an area on the map to be correctly labeled most of the time. The Kappa value for this assessment is greater than 0.80 representing a strong agreement and statistically significant GIS representation of tree canopy in Austin.

Although accuracy is high for the draft 2010 tree canopy data, a few limitations exist.

- Visual observation of the data shows accuracy decreases as distance increases from the center of tree canopy polygons. In other words, edge tree canopy is often not accurately represented, generally speaking.
- Larger stands of trees generally show higher accuracy compared to smaller stands of trees or individual trees.
- Variations in accuracy appear in different species of woody vegetation (i.e. mesquite vs. live oak); this is also observed in forest stands with varied age.

References

Congalton, R. G., & Green, K. (1999). *Assessing the accuracy of remotely sensed data: Principles and practices*. Taylor & Francis Group, Boca Raton.

Davey Resource Group. Prepared for City of Port Angeles. (2011). Urban tree canopy assessment. Retrieved January 14, 2014, from: http://www.cityofpa.us/PDFs/PermitsPlanning/UrbanTrees.pdf

Davey Resource Group. Prepared for City of Tukwila. (2012). Urban tree canopy assessment. Retrieved January 14, 2014, from: http://www.tukwilawa.gov/dcd/trees/Mar13_meeting/December%202012%20Final%20UTC%20Rpt.pdf

Davey Resource Group. Prepared for Indiana Department of Natural Resources Division of Forestry Community and Urban Forestry. (2011). Assessing and addressing Indiana urban tree canopy. Retrieved January 14, 2014, from: http://iufc.org/Final_IN_UTC_Summary_Report.pdf

Jensen, J. R. (1996). *Introductory Digital Image Processing: A Remote Sensing Perspective* (Second edition). Prentice Hall, Inc., Upper Saddle River, New Jersey.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33 (1): 159–174.

Lillesand, T., Kiefer, R. W., & Chipman, J. (2004). Digital image processing. In *Remote Sensing and Image Interpretation* pp. 550-577. John Wiley & Sons, Inc.

Skirvin, S. M., Kepner, W. G., Marsh, S. E., Drake, S. E., Maingi, J. K., Edmonds, C. M., Watts, C. J., & Williams, D. R. (2004). Assessing the accuracy of satellite-derived land-cover classification using historical aerial photography, digital orthophoto quadrangles, and airborne video data. In *Remote Sensing and GIS Accuracy Assessment* pp. 119-120. CRC Press. Retrieved January 14, 2014, from: http://www.epa.gov/esd/land-sci/pdf/l_1443_c09.pdf

Texas Trees Foundation. (2010). *The Dallas GIS roadmap model for urban tree planning and planting: The potential urban forest of Dallas, Texas*. Retrieved January 14, 2014, from: http://www.texastrees.org/cms/wp-content/uploads/2010/07/The-Dallas-GIS-Roadmap-Report_Final1.pdf

USDA, Forest Service website. (2009). Retrieved January 14, 2014, from: http://www.nrs.fs.fed.us/urban/utc/about/

USDA, National Agriculture Imagery Program. (2013). Information sheet. Retrieved January 14, 2014, from: http://www.fsa.usda.gov/Internet/FSA_File/naip_info_sheet_2013.pdf