Trees of green: constructing panels of tree canopy from aerial imagery

Alba Miñano-Mañero[*†]
CEMFI

February 2023

ABSTRACT: This paper develops a fully-automated workflow for constructing panels of tree canopy from high-resolution multispectral imagery with limited near-infrared (NIR) training data. The proposed workflow utilizes the tree-pixel detection algorithm developed by Yang, Wu, Praun, and Ma (2009) and Bosch (2020) on a large set of U.S. urban areas but modifies it by creating automatic ground-truth masks through various visual graphics techniques that leverage modern high-resolution NIR data. By matching colors across different imagery periods, the workflow predicts tree presence in older images without NIR data, using the recent images with NIR data. Using a subset of cities that represent the different U.S. climate regions, I quantify the effectiveness of the workflow by implementing the algorithm without pre-processing in the creation of ground-truth masks, without equalizing colors across periods, and using a universal model for all areas. The comparison shows that my workflow is the option that leads to better results in terms of accuracy, recall, and precision.

Key words: aerial imagery, tree detection, near-infrared light, panel data

[*]I am grateful to my advisor, Diego Puga, for all his guidance and help. I am likewise indebted to Cay Chaves for all his help automatizing the imagery download process.I gratefully acknowledge funding from Spain's Agencia Estatal de Investigación (MCIN/AEI/10.13039/501100011033) through its María de Maeztu Units of Excellence program (grant CEX2020-001104-M).

[†]CEMFI, Casado del Alisal 5, 28014 Madrid, Spain (e-mail: alba.minano@cemfi.edu.es; website: https://albaminanomanero.github.io).

1. Introduction

The distribution of urban trees shapes the environmental and social fabric of modern cities. Their benefits span from providing aesthetic value to numerous ecological and health advantages, like reducing heat-island effects by providing shade and reducing storm-water runoff, inducing energy-savings, sheltering from the wind, improving mental well-being and reducing crime (Morales (1980); Livesley, McPherson, and Calfapietra (2016); McPherson, van Doorn, and de Goede (2016); Reid, Clougherty, Shmool, and Kubzansky (2017); Shepley, Sachs, Sadatsafavi, Fournier, and Peditto (2019); Jones (2021)). Studying and monitoring the distribution of trees within a city is key to assessing and maximizing those benefits, as well as promoting equitable access to natural amenities and addressing environmental injustices related to the unequal provision of tree coverage (Schwarz, Fragkias, Boone, Zhou, McHale, Grove, O'Neil-Dunne, McFadden, Buckley, Childers, Ogden, Pincetl, Pataki, Whitmer, and Cadenasso, 2015). Authorities have typically tracked the urban canopy with the conduction of manual tree inventories with the location, species, and condition of public trees. However, the recent advances in multispectral imagery and machine learning can allow us to obtain similarly precise data replacing the labor intensiveness of conducting manual tree inventories.

This study presents a new pipeline that utilizes high-resolution aerial imagery, visual graphics techniques, and machine learning pixel-classification algorithms to automatize the generation of urban tree coverage panel data. The workflow is capable of predicting the presence of tree coverage at the pixel-level, even when there is limited data available to create ground-truth masks for training areas. One of the main strengths of this method is its potential for widespread application, given that multi-spectral aerial imagery is publicly available in multiple periods and geographic locations. Moreover, this paper has implemented the proposed workflow to a geographic scale that covers multiple urban areas of the United States in two time periods, representing a much more extensive coverage than other similar proposed algorithms.

Implementing any machine-learning algorithm for tree detection requires having available highly geographically precise training data. Typically, it is obtained from city inventories with limited geographic coverage (Beery, Wu, Edwards, Pavetic, Majewski, Mukherjee, Chan, Morgan, Rathod, and Huang, 2022); using training data annotated by hand with bounding-box annotations (Wegner, Branson, Hall, Schindler, and Perona (2016); Weinstein, Marconi, Aubry-Kientz, Vincent, Senyondo, and White (2020); Weinstein, Graves, Marconi, Singh, Zare, Stewart, Bohlman, and White (2021)) or point annotations (i.e., Chen and Shang (2022)) or by-hand point annotations combined with inventories as in Ventura, Honsberger, Gonsalves, Rice, Pawlak, Love, Han, Nguyen, Sugano, Doremus, et al. (2022). The hand annotation process is not only costly, but it can also lead to inaccuracies when the algorithms are designed for pixel-level segmentation and not for object detection. The alternatives then involve using light detection and ranging (LiDAR) point-cloud data, since it captures high-resolution ground-elevation data, or near-infrared light, which captures vegetation photosynthesis. However, the former is expensive and geographically limited, and the latter is only available at high-resolution and

large geographic scales recently. This paper develops a way to use near-infrared for just one period to predict the presence of trees at a pixel resolution for periods in which this data is unavailable.

Although there is a growing body of literature implementing the most modern visual graphic techniques and machine learning algorithms to detect the presence of trees and their coverage, they focus less on the transferability of these algorithms across periods. One of the reasons is the difficulty of homogenizing images. Also, the time perspective is less developed due to the difficulties of obtaining ground truth data for multiple periods. This work fills in this gap by developing a workflow that relies on the algorithm developed by Yang et al. (2009) and Bosch (2020), and that predicts tree coverage in multiple periods with only one period of training data. To do so, it exploits near-infrared (NIR) data to fully automatize the creation of ground-truth masks just for one time period and use them for periods when NIR light is unavailable.

The main contribution of this paper is developing an approach to construct panels of tree coverage with high-resolution aerial imagery for areas of interest with periods without NIR data. To do so, the first step consists of equalizing the colors of the non-NIR images to match the colors of the same geographic area for the period in which the NIR data is available. This step is necessary as the Yang et al. (2009), and Bosch (2020) (YWPM&B henceforth) method is based exclusively on colors. Hence, using a trained model to predict areas with very different colors would hinder transferability. For the period without NIR data, once the training areas have been established for the first period following YWPM&B methodology, the tree-pixels to create the ground-truth masks are labeled using iterative thresholding on the images for the same area of the period with near-infrared data. The YWPM&B algorithm is then trained using these images and implemented to predict images in the other period (i.e., the one without NIR data). For periods in which NIR data is available, the workflow simply automatizes the creation of pixel-level accurate ground-truth masks.

In a similar vein as Ventura et al. (2022) and Beery et al. (2022), but unlike Weinstein, Marconi, Bohlman, Zare, and White (2019), which focuses on large-scale tree detection in natural forests, I implement the workflow in an extensive collection of urban areas across 36 different Metropolitan Statistical Areas in the United States. However, in contrast to Ventura et al. (2022), this paper identifies urban tree cover (pixel-level detection) rather than detecting individual trees (object detection). The detection of coverage matters because changes over time in tree canopy entail changes both in the extensive and areal margin, which implies that to approximate its evolution the employed algorithm needs to be able to detect new trees but also increased leaf or crown area of trees already present, as it is positively associated with the environmental benefits of trees (Pretzsch, Biber, Uhl, Dahlhausen, Rötzer, Caldentey, Koike, van Con, Chavanne, Seifert, du Toit, Farnden, and Pauleit, 2015). While object detection can accurately capture changes in the amount of trees, it is unlikely to detect other changes in coverage that pixel-level detection can capture. Furthermore, after detecting pixel-level canopy, stand-alone trees or groups of trees can be obtained with some simple GIS processes. Moreover, although Ventura et al. (2022) emphasize the large scale of their analysis, this paper

goes beyond their scope by considering multiple urban areas in two periods of time spread across the United States and not a single state. Back of the envelope calculations imply that the areas under consideration in this paper cover 0.1% of the U.S. area, and 7% of its urban area, while the numbers for Ventura et al. (2022) are 0.0004% and 0.02% respectively[1]. Moreover, rather than applying a model estimated from a specific geographic area, this paper provides trained models for two periods in each study area. Unlike Beery et al. (2022), this workflow does not require mapping tree census data to imagery for model training, which increases its applicability to areas without tree census data. In addition, this research differs from Beery et al. (2022)'s approach by focusing on detecting tree canopies rather than automated species identification.

The development of deep and convolution neural networks has allowed the emergence of the literature focused on tree delineation without depending on using expert-engineered features. Although some recent test-training datasets have emerged using these methodologies, the technical expertise required to manipulate these algorithms and the need for extensive training data and computational resources limits their applicability (Weinstein et al., 2020). This paper, instead, relies on YWPM&B's detection algorithm that requires no fine-tuning beyond providing accurate ground-truth masks and produces accurate results. Precisely, the automatic generation of these ground-truth masks and their transferability across periods developed by this workflow allows obtaining maps of tree coverage and statistics for geographic units over time, without the need for specific expertise or computational resources.

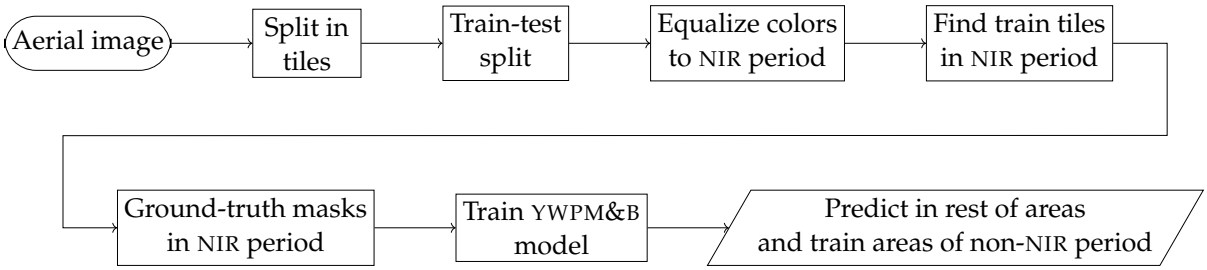
To quantify the relevance of using the proposed workflow, this paper trains and predicts the model under various scenarios in a subset of areas representing seven out of the nine climate zones in the United States. To assess the individual relevance of each step of the pipeline, the experiments involve removing these steps one at a time. Each experiment uses a 5% and 10% sample of tiles to guarantee representative results. The analysis of the confusion matrices indicates that equalizing colors and fitting models for each area are crucial for obtaining accurate predictions, although there is some heterogeneity across cities.

This paper contributes to the remote sensing literature by bridging the gap between image processing and machine learning methods for tree detection. To achieve this, the proposed workflow utilizes image processing techniques to generate ground-truth masks that train pixel-level artificial intelligence algorithms[2]. While a range of vegetation indexes exist to establish the feature selection, this paper combines the visible and non-visible channels of multi-spectral imagery to compute the standard Normalized Difference Vegetation Index (NDVI) that can separate tree/not-tree pixels. After some pre-processing to smooth the image, background, and foreground pixels are defined by thresholding the NDVI histogram in each area. Strict image-processing methods would perform this operation everywhere to classify pixels as a

[1]The area covered is computed as the number of tiles multiplied by the area of each tile, which in the case of this paper is 512^2 sq.m. and for Ventura et al. (2022), $(256*60)^2$ sq. cm

[2]To see a recent survey on different remote-sensing methods for tree detection see Hanapi, Shukor, and Johari (2019)

Figure 1: Workflow



Notes: Figure describes the workflow for period where there is no near-infrared data available.

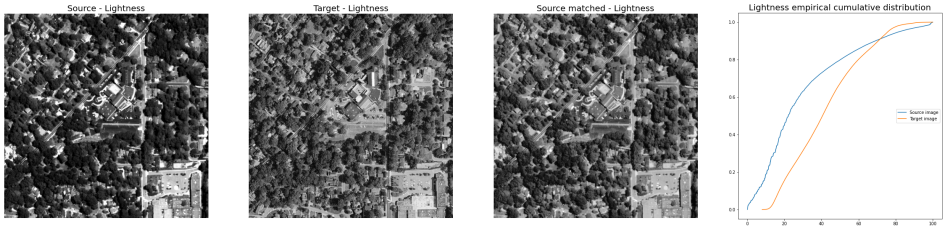
tree-not tree, thus being restricted by the availability of multi-spectral imagery with the needed channels to compute the index. However, because the image segmentation methodologies achieve highly accurate results (for instance, a 90% detection rate in Srestasathiern and Rakwatin (2014)), this paper employs them to label the training areas, fit the YWPM&B model and predict it in images that do not contain all the needed channels to perform image segmentation.

2. Methodology

The methodology used in this paper to construct panels of tree canopy consists of three distinct elements. Firstly, it involves the equalization of colors across different periods. This step is important for ensuring that the color variations in the images do not affect the accuracy of the subsequent analysis. Secondly, ground-truth masks are created for the training area to provide a reference for the machine learning algorithm. Finally, a pixel-classifier algorithm is estimated using YWPM&B. While this algorithm provides a crucial component of the methodology, the first two steps and their integration into an automated workflow are unique contributions of this paper.

Figure 1 illustrates the workflow for situations without near-infrared (NIR) training data. Appendix Figure 5.1.1 shows the aerial image of Manhattan as example of the images used. The process starts by splitting these aerial red-green-blue (RGB) images into 512×512 pixel tiles. 1% of the tiles are selected as training areas with YWPM&B's proposed train-test split method. To ensure the training data is representative, the tiles are split into clusters with k-means clusters based on the tiles' summary characteristics (i.e., statistical information of texture and colors). As many clusters as training areas are created, and for each cluster, the cluster's centroid tile is added to the training set. Next, the colors of the tiles are matched to those of the tiles in the period with NIR data, and the training areas are replaced with images from the same area in the NIR period. Ground-truth masks are created for the training areas, and the YWPM&B classifier is trained. Finally, the trained classifier is used to predict all areas, including the areas that were set as training in the non-NIR period. The workflow in the NIR period is essentially the same, but without the equalizing color and training tiles substitution steps.

Figure 2: Histogram matching on the lightness channel



Notes: Figure illustrates the histogram matching process. The image on the left is the image whose histogram has to be matched to the image on the middle. The third image is the matched one after implementing the histogram matching, and the last one represents the histograms of the source and target images.

To ensure that images have consistent colors across different time periods, this paper employs color equalization techniques. When dealing with images taken from a similar viewpoint and with similar sensors, histogram matching is an effective way to standardize colors (Shapira, Avidan, and Hel-Or, 2013). Image histograms represent the pixel intensity distributions, and their manipulation serves various purposes, such as segmentation and contrast enhancement (Nikolova, Wen, and Chan, 2013). Histogram matching involves mathematically transforming the cumulative distribution function of the source variable to match that of the target variable. In image processing, where pixel values are typically discrete and bounded, histogram matching involves creating a lookup table for each value. Figure 2 exemplifies the process. However, while histogram matching, and the rest of histogram manipulations, are straightforward for gray-scale images with only one channel, the application for colored images is subject to certain limitations since, unlike gray-scale images, colored images have multiple channels to match.

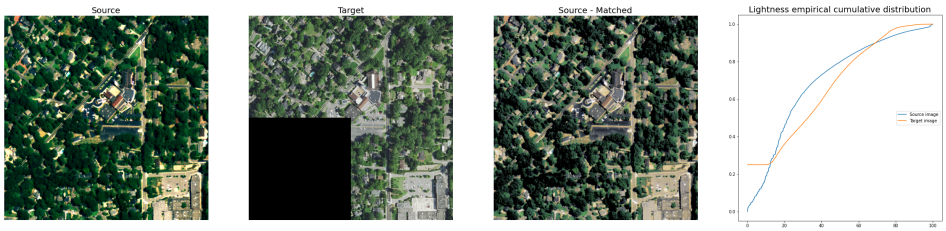
In general, it is unwise to match each of the color channels independently as the resulting colors will be erroneous (Gonzalez and Woods, 2018), and in the case of images in the red-green-blue (RGB) color space –as the ones this paper uses– the results will be unpredictable due to the correlation between channels and the lack of perceptual uniformity (Grundland and Dodgson, 2005). Rather than using the computationally intensive joint-histogram equalization, this paper transforms the images to the $L^*a^*b^*$ color space to match the histogram of each of these new channels independently. The $L^*a^*b^*$ color space is a perceptually uniform color space (i.e., perceived difference between two colors is proportional to the model one) where the channels represent perceptual lightness (L^*) and the unique colors of human vision according to the opponent color model (red-green in the a^*, and blue-yellow in b^*). Because of the perceptual uniformity and opponent nature of the $L^*a^*b^*$ color space, their channels are less statistically correlated and can be matched independently without resulting in erroneous colors[3]. Visual inspection of the resulting matched images using different color spaces (i.e., HSV, LCH) also showed that the results using $L^*a^*b^*$ outperformed the others. Thus, the strategy for color equalization implies converting the images from RGB to $L^*a^*b^*$ color space, matching the

[3]Other papers have also found that the $L^*a^*b^*$color space outperforms the rest for color matching as Grundland and Dodgson (2005) and Sunkavalli, Johnson, Matusik, and Pfister (2010)

Figure 3: *L*a*b* histogram color matching

Notes: Figure illustrates the result of the color matching. The left-most image is the source image that is matched to the middle image. The right-most image is the resulting matched image.

Figure 4: *L*a*b* histogram color matching

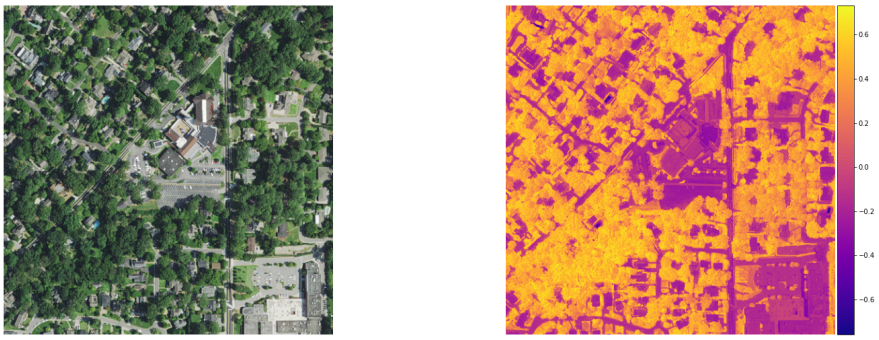


Notes: Figure illustrates an extreme scenario in which there are extreme divergences between source and reference images.

histograms of the *L*a*b** channels independently, and reconverting them back to RGB to be able to implement the rest of the algorithms. Figure 3 provides an example of the result of the procedure.

The paper implements the technique in a localized way matching the histograms of the same 512x512 $1m^2$ tile in the two time periods. This is needed because histogram matching requires similar features so that the difference in histograms is attributable to changes in lighting and coloring conditions and not a drastic difference in the object image composition. To see this consider an extreme case, as shown in Figure 4 in which the reference image has a portion with all black pixels, then the target L^* CDF turns flat on zero –which corresponds to black– on the percentage of black pixels (25% in this case). Then, once the source CDF is matched, the 25% source pixels in the lowest part of the lightness distribution would transform into black. While this is an extreme case, similar coloring effects will happen when the histograms differ substantially.

Figure 5: NDVI

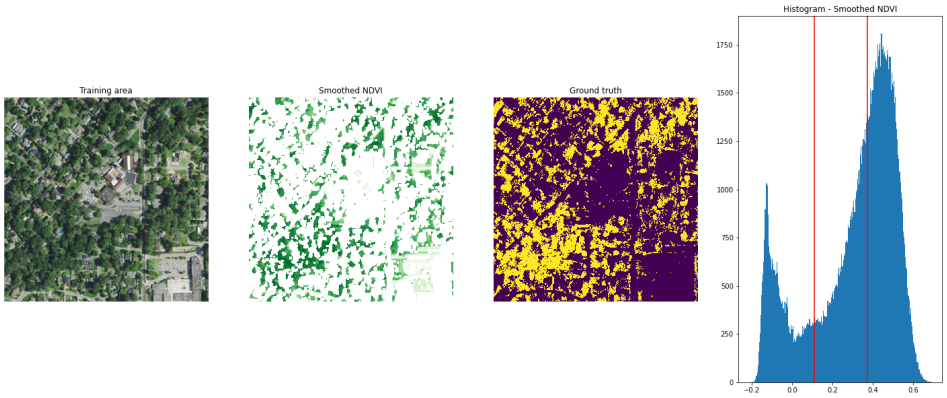


Notes: Figure depicts the NDVI of the example training area.

Creating ground-truth masks leverages the availability of high-resolution aerial imagery containing the two spectral bands typically used in the remote sensing of vegetation: red (R) and near-infrared (NIR) light, whose wavelengths lie just outside the range of human vision. Alive foliage appears green to the human eye because the chlorophyll pigment absorbs most of the visible wavelengths for photosynthesis while reflecting green radiation. In contrast, the cell structure of leaves reflects a significant portion of the NIR radiation received back into space. Hence, due to chlorophyll pigment, healthy vegetation will concentrate on the lower values of the red band and seem dark, but in the NIR band, the leaf cell structure results in high reflectance. Traditionally, this behavior is summarized via the normalized difference vegetation index (NDVI): $\frac{\text{NIR}-R}{\text{NIR}+R}$, with values ranging from -1 to 1. While vegetation will always have positive values, the higher the value, the greener, more alive, and denser the vegetation; negative values are typically associated with water, and low values are typically associated with areas of soil, rock, sand, and man-made structures. Moreover, because soil spectral reflectance is not different in the R and NIR bands, the NDVI can be used to separate vegetation from the background (Karnieli, Agam, Pinker, Anderson, Imhoff, Gutman, Panov, and Goldberg, 2010). An example area with its NDVI can be seen in Figure 5.

As there is no universal value that can distinguish among different classes, since the NDVI is affected by various factors such as soil type, brightness, crop type, and growth stage (Xue and Su, 2017), this paper implements nested image thresholding to find an NDVI separating value for each particular training tile. Thresholding is done using Otsu's method (Otsu, 1979), a widespread image segmentation technique that consists of finding the value that maximizes between-class variances for the histograms. In the first round, thresholding allows to separate the shadow from sunlit pixels and then, in the second round, to detect vegetation in sunlit pixels, similarly to Otsu, Pla, Duane, Cardil, and Brotons (2019). As shadows are typically found on the first valley of the NIR histogram (Adeline, Chen, Briottet, Pang, and Paparoditis, 2013), Otsu's thresholding is applied on the NIR band smoothed with Gaussian blur on a 3-by-3

Figure 6: Creation of ground-truth masks

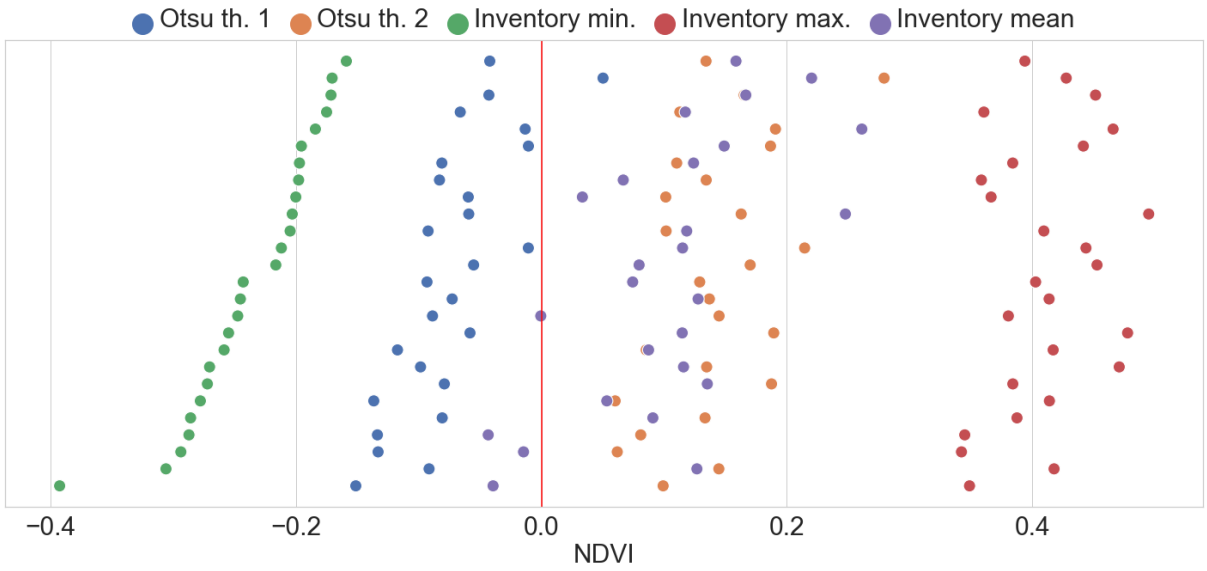


Notes: Figure illustrates the thresholding algorithm in the example training area. The first image is the training area; the second image is the smoothed NDVI after having removed shadows; the third image represents the final ground-truth masks where yellow pixels are the ones that are labeled as true pixels; the histogram shows the threshold values for the smoothed NDVI image histogram.

window with a standard deviation of 5 to reduce noise. Then, keeping only the pixels that are above the threshold, the NDVI is computed. After applying the same Gaussian blur filter to the NDVI, Otsu's method is implemented to segment the image into three categories rather than two. The intuition for using three classes is that since most urban areas exhibit mixed features with different NDVIs, it guarantees that the highest class will correspond to the densest, greenest, and most alive vegetation. All pixels whose NDVI falls above the second threshold are labeled as true tree pixels in the ground truth mask. The process is illustrated in Figure 6.

To assess how well this ground-truth masks map to actual trees, New York City provides an ideal scenario as it has aerial images for 2015, the year in which the city also conducted its third street tree census. Sampling the XY coordinates of the alive trees from the census to the ground-truth masks of the training areas shows that approximately 40% of the census trees fall in a tree-labeled pixel. As NDVI values are the basis for creating the ground-truth tiles, understanding these discrepancies requires sampling the NDVI values at tree coordinates and comparing them to the thresholds. Figure 7 depicts the relationship between thresholds and NDVI sampled distribution at tree coordinates for the training areas. To summarize the distribution of NDVI values for tree census locations of each tile, the Figure shows the minimum, the maximum, and the mean NDVI and compares them to the two Otsu's thresholds. All points along the same horizontal line represent the values for the same training tile. The missing trees would be the ones whose NDVI lies between the inventory minimum and Otsu's second threshold. The first thing to notice from this figure is that the minimum sampled NDVI for census trees is (1) always negative, and (2) always below the first threshold. The situation represents that the missing tree coordinates correspond, according to its NDVI, to mand-made structures and, according to the segmentation, the lowest category. The implication is that

Figure 7: Comparison of NDVI thresholds and tree sampled NDVI



Notes: Figure shows the location of the NDVI sampled distribution for tree XY coordinates of the city inventory and the estimated thresholds for the training tiles of New York City. All points on the same horizontal line correspond to the same tile.

the trees that are missing would never be seen using the NDVI. The potential reasons behind this relate to the angle the image was taken from, which can imply the trees are covered with buildings or the coordinates might not correspond to the crown centroid. Notice, however, that for tiles with positive mean inventory NDVI, the second threshold tends to be similar and is always below the maximum, implying it successfully captures the top of the distribution.

After creating the ground-truth masks, the YWPM&B pixel-classifier is trained. The classifier is based on visual features that represent the most distinctive properties of trees. The features represent colors in RGB and *L*a*b* color spaces, texture applying Gaussian derivative filters to the L^* channel and entropy, which tends to be higher in trees than man-made structures, on the L^* channel. Then, adaptive boosting is used to train a strong classifier based on weak classifiers constructed from these features. Then the model is used to predict the presence of trees in each pixel of the rest of images and the predictions are refined to avoid stand-alone tree pixels.

Assessment metrics

The performance of the workflow is assessed with three different standard metrics in machine learning: accuracy, recall, and precision. Given a confusion matrix (M) with cells containing the share of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for an area of interest, Equations 1 to 3 define the measures:

$$Accuracy = \text{tr}(M) = TF + TN \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

The usefulness of the measures lies in the fact that they provide a quantitative assessment of how well the algorithm is predicting. In particular, accuracy measures the proportion of pixels correctly classified across the tree and not-tree categories. Recall, instead, quantifies the model's ability to label true tree pixels as such. Precision, finally, measures how many of the tree-labeled pixels are real trees.

Since obtaining a confusion matrix for tiles requires observing the model predictions as well as the ground truth of that location, evaluating the performance of the models is only feasible in the period with NIR data when we can construct the true labels for pixels as described in Section 2. Hence, after training and predicting the model for each urban area under study, the performance metrics are computed using a 5% random sample of tiles for each study area. When experimenting with the model in a subset of cities, the assessments samples a 5% and a 10% of tiles, to decrease the likelihood that outliers in prediction are determining the differences.

Data

Implementing the methodology described uses high-resolution aerial imagery in two time periods from the National Agriculture Imagery Product (NAIP) conducted by the United States Department of Agriculture. The program started in 2003 and has been re-conducted every three years since 2009. Images are always taken during the agricultural growing sessions, ensuring leaf-on conditions for trees, which allows for their detection from the sky.

While during the first rounds, images were natural color (red-green-blue) images, recent periods also contain the near-infrared band. In general, images for both periods are at a $1m^2$ resolution, but for some states, in the second period, it increased to 0.6. In such cases, the image with the highest resolution is resampled to match the resolution of the other. In total, the study areas contain 51,414 512×512 tiles for each period, which would account for 0.1% of the US area and a 7% of the urban US area. Table 1 shows the urban areas with the corresponding number of tiles and the two years for which the algorithm has already been implemented.

Of the cities shown in Table 1, the tests of the algorithm are performed on a subset of those, aiming to represent the different climate zones of the US, but being less computationally intensive by using a smaller sample. The selected cities are the following: The Five Boroughs of New York City (Northeast zone); Flint (Upper Midwest); Akron (Ohio Valley); Birmingham (Southeast); New Orleans (South); Seattle (Northwest); San Francisco (West).

3. Results

This section presents the results of implementing the algorithm discussed in Section 2 on the study areas. Besides, to evaluate the effectiveness of the proposed workflow, a series of experiments are performed in which each of the workflow steps is modified to analyze its

Table 1: Distribution of tiles per city

City	Tiles	City	Tiles	City	Tiles
Akron, OH (2004-2015)	1,305	Columbus, OH (2004-2015)	1,024	Queens, NY (2006-2015)	1,972
Atlanta, GA (2007-2015)	1,557	Dayton, OH (2004-2015)	504	Richmond, VA (2003-2015)	360
Baltimore, MD (2005-2015)	1,064	Detroit, MI (2005-2014)	5,368	Rochester, NY(2006-2015)	775
Birmingham, AL (2006-2015)	874	Flint, MI (2005-2014)	936	San Francisco, CA (2005-2014)	550
Boston, MA (2003-2014)	1,085	Kansas, MO (2007-2015)	647	Seattle, WA (2006-2015)	1,263
Bronx, NY (2006-2015)	572	Los Angeles, CA (2005-2014)	12,536	Somerville, MA (2003-2014)	16
Brooklyn, NY (2006-2015)	1,094	Manhattan, NY (2006-2015)	646	St. Louis, MO (2007-2015)	2,560
Buffalo, NY (2006-2015)	511	Milwaukee,WI (2005-2015)	844	Staten Island, NY (2006-2015)	1,089
Cambridge, MA (2003-2014)	104	Nashville, TN (2006-2014)	442	Syracuse, NY (2006-2015)	462
Camden, NJ (2006-2015)	154	New Haven, CT (2006-2014)	420	Toledo, OH (2004-2015)	872
Chicago, IL (2007-2015)	3,448	New Orleans, LA (2007-2015)	650	Trenton, NJ (2006-2015)	165
Cleveland, OH (2004-2015)	4,034	Oakland, CA (2005-2014)	1,051	Westchester, NY (2006-2015)	460

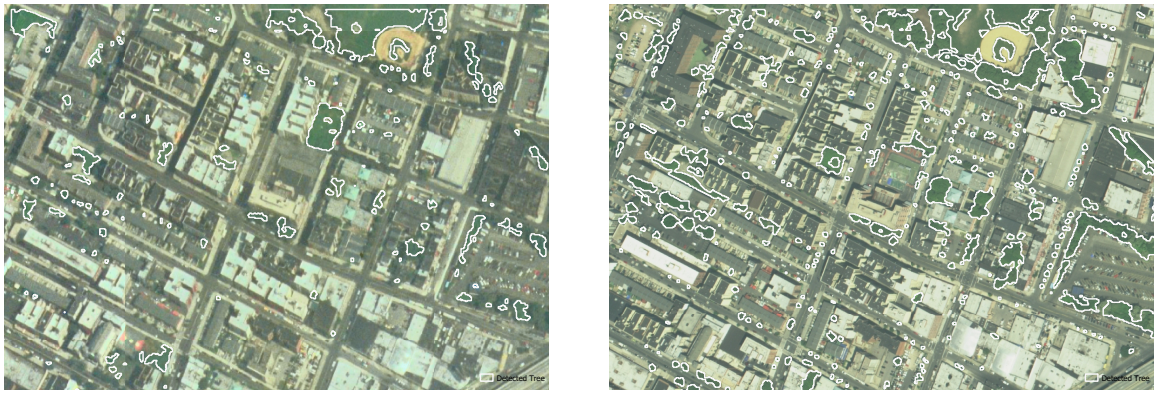
Notes: This table shows the number of 512× 512 tiles per city and the two years for which the algorithm has been implemented.

impact on accuracy, recall, and precision. These experiments provide valuable insights into the performance of the algorithm and help identify the most critical factors that contribute to achieving accurate and reliable results.

Overall the algorithm successfully identifies and maps tree canopy cover in the study areas, as Figure 8 shows. The figure displays the model prediction for a Manhattan area in the two time periods showing that the algorithm captures changes over time in tree coverage for a location. Accuracy, recall, and precision values were consistently high across all study areas, indicating the robustness and reliability of the proposed workflow. Figure 9 shows the three assessment metrics for the cities on which the model has been implemented, and Appendix Table 5.1.1 shows the corresponding numerical values. On average, accuracy values were consistently the highest, with an average of around 0.9, and a minimum of 0.8 in Birmingham. This implies that, on average, approximately 90% of the pixels are classified as tree/not tree correctly. Recall rates exhibit more variation: while for some cities the model predicts 80% of true tree pixels to be tree pixels, for Oakland and LA, recall rates fall around 0.5 (i.e., half of the truth tree pixels are not labeled correctly). This variation in recall rates may be attributed to the tree cover in Oakland and LA being more diverse and complex, including areas of urban tree forest, which hampers the model's ability to identify tree pixels. However, the average recall rate for the model is 0.77, implying that, on average, the model can correctly predict the presence of trees. However, the precision rate, although having a similar average as recall, exhibits less variation and is no lower than 0.6 (Rochester), implying that most tree-labeled pixels correspond to actual trees.

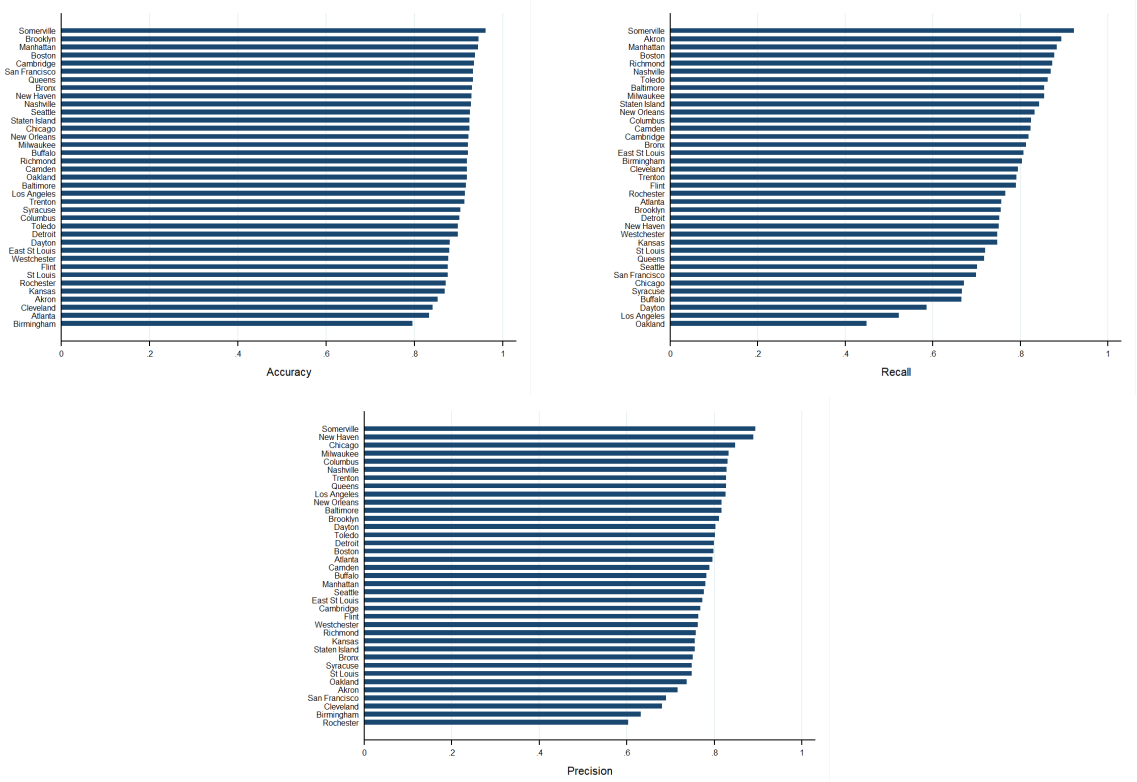
In summary, the workflow successfully identifies tree canopy in the study areas with high accuracy, recall, and precision. While recall rates exhibit more variation due to the complexity of tree cover in some cities, the average recall rate is still high, and the precision rate indicates that most tree-labeled pixels correspond to actual trees. Overall, the algorithm is robust and reliable in capturing changes over time in tree coverage for a location.

Figure 8: Example of detected canopy in Manhattan in 2006 (left) and 2015 (right)



Notes: Figure shows the predicted tree pixels (in white) for an area of Manhattan in two periods of time, 2006 (left) and 2015 (right)

Figure 9: Assessment metrics of baseline model: accuracy (left), recall (right), precision (bottom)



Notes: Figure shows the accuracy, recall and precision rates estimated according to Equations 1-3 using the baseline model describe in Section 2.

Experimentation

To evaluate the effectiveness of the proposed algorithm for mapping tree canopy cover, several experiments were conducted modifying the different steps described in Figure 1. These experiments aimed to investigate the impact of the ground-truth map creation, the lack of color equalization, and the usage of a universal model. This subsection presents the results of the experiments and discusses their implications for improving the accuracy and robustness of the algorithm. Given that the results may vary depending on the specific study area, the experiments are performed in a subset of urban areas that ought to represent most of the different climate zones of the US.

Modifying ground-truth mask creation

The first experiment involves the modification of the ground-truth mask creation. The proposed algorithm generates label pixels in training data in a series of steps that aim at reducing noise in the NDVI distribution for a tile so that Otsu's thresholds can accurately capture the valleys and picks in the histogram. To do so, it started removing shadowed pixels by eliminating those for which the NIR reflectance was below the first threshold of the smoothed NIR band. Then, it obtained the NDVI with the remaining sunlit pixels and smoothed it before applying the multi-Otsu's thresholding. Finally, every pixel whose NDVI was above the second threshold obtained in this last step was labeled as a tree pixel.

The relevance of this procedure is evaluated with two different experiments. The first experiment uses only the multi-Otsu thresholding on the NDVI to generate the ground-truth masks. The meaning is that, for each training tile, all pixels whose NDVI falls above the second threshold are classified as trees. The second one, however, compares the predictions of using this simple methodology to the ground-truth data using all the pre-processing steps (shadow removal and smoothing). Notice that these two experiments capture two differences in performance: (1) in general, how well does YWPM&B procedure perform when using different training data, and (2) how well does it work when training data does not fully align with the ground truth data.

Table 3 shows the results of the two experiments. The marginal differences in accuracy between the experiment and full processing imply the algorithm can produce truthful predictions when estimated with different training data. Comparing the first experiment to the original one indicates that while full pre-processing is associated with the higher recall, the experiment is associated with higher precision. The implication is that the proposed model is better at capturing tree cases, although it may produce more false positives. The experiment, however, tends to label more tree pixels as such but may predict more false negatives. The explanation is that the complete model eliminates pixels with NIR in the bottom category, and pixels that pass the first cut-off are, therefore, more likely to be labeled as trees reducing false negatives (i.e., trees not detected as trees). However, it may also increase the number of false positives (i.e.,

Table 2: Experimentation: modifying ground-truth masks

	Exp. 1: NDVI Threshold			Exp. 2: NDVI threshold vs. full processing		
City	Accuracy	Recall	Precision	Accuracy	Recall	Precision
Akron	0.87	0.85	0.82	0.85	0.89	0.70
Birmingham	0.87	0.80	0.82	0.81	0.92	0.63
Bronx	0.92	0.91	0.81	0.90	0.91	0.66
Brooklyn	0.94	0.81	0.85	0.93	0.82	0.68
Flint	0.89	0.85	0.87	0.84	0.95	0.65
Manhattan	0.95	0.81	0.88	0.95	0.83	0.78
New Orleans	0.92	0.89	0.87	0.89	0.95	0.69
Queens	0.94	0.82	0.87	0.93	0.88	0.72
San Francisco	0.91	0.46	0.70	0.94	0.52	0.69
Seattle	0.82	0.53	0.81	0.84	0.66	0.51
Staten Island	0.93	0.85	0.86	0.92	0.92	0.71
Westchester	0.88	0.87	0.80	0.84	0.92	0.63

Notes: Table displays the performance results of running the algorithm using only double thresholding on NDVI (Exp. 1) and training the algorithm defining with only NDVI thresholding but comparing to ground-truth data using all the full pre-processing (Exp 2.) on a 5% of tiles.

non-trees identified as trees) because some pixels with low NDVI values may still be classified as trees due to the Gaussian blurs.

While differences between both algorithms tend to be minimal, they enlarge in certain cities like San Francisco and Seattle, where recall rates would fall from 70% to around 50% with the experimentation, which may be related to the particular layout of these areas. Comparing the performance using a 10% sample of tiles, shown in Appendix Tables 5.1.2 and 5.1.3, indicates similar differences, with the full-processing algorithm associated with higher recall and lower precision. However, while the performance of the full processing is stable, recall rates of the experimentation have high variability and fall to a minimum of 0.2 in San Francisco. The implication is that the experiment is not as robust to data variation as the complete pre-processing this paper implements to create the ground-truth masks.

The second experiment, which creates training data directly thresholding the NDVI, but compares the prediction of this model to the creation of ground-truth masks with all the processing steps, allows assessing how well YWPM&B model performs when trained with the *wrong* data, meaning data that does not correspond 100% to the ground-truth masks. Comparing the results with both the 5% and 10 % sample (Appendix Table 5.1.3) shows that the second experiment is associated with higher recall rates but lower precision than the correct model. Behind this is the fact that training the model with direct thresholding leads to a model that automatically labels more pixels as trees than it should -according to original ground-truth masks- in this way, while the model reduces false negatives by classifying more pixels as trees, which increases recall, it also increases the number of false positives, decreasing precision. Thus the training data must represent the ground-truth reality to achieve models with good performance.

Table 3: Experimentation: not-equalizing colors

	Using a 5% sample			Using a 10% sample		
City	Accuracy	Recall	Precision	Accuracy	Recall	Precision
Akron	0.79	0.79	0.64	0.79	0.79	0.63
Birmingham	0.78	0.93	0.60	0.82	0.92	0.65
Bronx	0.86	0.00	0.05	0.86	0.00	0.37
Brooklyn	0.90	0.00	-	0.92	0.00	0.00
Flint	0.85	0.56	0.63	0.84	0.60	0.64
Manhattan	0.88	0.00	-	0.87	0.00	0.97
New Orleans	0.88	0.50	0.82	0.90	0.56	0.84
Queens	0.85	0.00	0.05	0.87	0.00	0.68
San Francisco	0.96	0.54	0.84	0.94	0.48	0.72
Seattle	0.85	0.14	0.24	0.85	0.21	0.26
Staten Island	0.82	0.00	-	0.86	0.00	-
Westchester	0.74	0.00	0.07	0.74	0.00	0.08

Notes: Table displays the performance results of training the model with tiles from a given color but implementing it in tiles whose colors have been equalized to other tiles' histograms using a 5% and a 10% of tiles.

Not equalizing colors

The second experiment aims to evaluate the impact of color normalization techniques on the transferability of models. The design is as follows: while the model is trained using the images with colors of the period with NIR data, and then the performance is evaluated on tiles of the city that have been color equalized to match the colors of the first period images with *L*a*b* histogram matching. This experiment is, in this way, able to determine the extent to which color normalization techniques are necessary for the transferability of the model.

Table 3 shows the performance metrics for the experiment using both a 5% and 10% sample of tiles. According to the results, one of the findings is across-city heterogeneity. While the relatively low magnitudes for recall and precision in cities outside New York are consistent with relatively worse-performing models, the values indicate the models fail in New York: in these areas, the model labels very few or no single pixels as trees. The implication is that the accuracy value in those areas is simply the percentage of pixels that are not-tree pixels, recall rates are zero, and precision values are extremely low or missing. Comparing the two samples the situation experiments marginal changes, except for some of the boroughs of New York, where precision rates explode while the accuracy and recalls remain substantially unchanged. The explanation for these changes is that with a larger sample, the models label a negligible share of true trees as a tree pixel, causing a substantial increase in the precision rate. Moreover, similar accuracy and recall rate for New York with a larger sample size suggests that the observed performance is not a result of biased sampling or localized poor-quality imagery.

Observing these two extreme situations implies there is a subset of cities, like New York, for which color equalization is an imperative requisite for transferability, while for others it

Table 4: Experimentation: using a universal model

	Using a 5% sample			Using a 10% sample		
City	Accuracy	Recall	Precision	Accuracy	Recall	Precision
Akron	0.72	0.08	0.89	0.71	0.08	0.93
Birmingham	0.84	0.92	0.63	0.80	0.88	0.60
Bronx	0.91	0.72	0.70	0.92	0.74	0.73
Brooklyn	0.92	0.66	0.70	0.92	0.66	0.68
Flint	0.72	0.00	0.02	0.74	0.00	0.02
Manhattan	0.95	0.76	0.77	0.92	0.79	0.69
New Orleans	0.86	0.92	0.62	0.87	0.93	0.66
Queens	0.91	0.53	0.77	0.92	0.57	0.75
San Francisco	0.90	0.61	0.50	0.91	0.63	0.50
Seattle	0.91	0.58	0.79	0.90	0.51	0.82
Staten Island	0.92	0.78	0.69	0.93	0.73	0.76
Westchester	0.82	0.67	0.69	0.83	0.60	0.69

Notes: Table displays the performance results of training a universal model using 1% of all tiles and using it to predict in each area, with a 5% and a 10% sample of tiles.

improves the predictions as the performance rates are always superior using the correct color model. Despite implementing a consistent color equalization process across all experimental areas, there were significant variations in their performance rates, indicating that the color changes over time differed across the regions. Overall, the results suggest that incorporating color equalization is a promising approach for improving the transferability of algorithms across time.

Universal model

The last experiment evaluates the transferability of a universal model across different urban areas. Its goal is to determine if a single model would accurately predict tree cover in diverse geographic regions or if individual models are necessary for each area. In order to evaluate this hypothesis, a universal model is trained using a 1% sample of tiles from all geographies, following the k-means cluster train-test split of the YWPM&B model. The universal model is then used to predict the presence of tree pixels in 5% and 10% of tiles from each area and compared to their ground truth masks. By comparing the performance of the universal model with that of individual models for each urban area, this experiment sheds light on the feasibility of using a universal model for tree cover prediction across diverse geographic regions.

Results of the experiment with the two different samples are presented in Table 4. As with the color experiments, using a universally trained model to predict tree canopy reveals two distinct scenarios. In cities such as Akron and Flint, the model predicts an insignificantly low proportion of true positive tree pixels, yielding recall rates that are nearly zero, and accuracy rates that stem from labeling all pixels as non-tree. For the rest of the cities, the model generates

different tree-not-tree predictions but with performance rates that are always worse than using individual models for each of them. This phenomenon can be attributed to the heterogeneity of the data across the different cities. When training individual models for each city, the model can capture the unique characteristics of that particular area. However, a universally trained model cannot account for the heterogeneity resulting in lower performance overall. In cases like the present one, where there are high levels of heterogeneity across areas, the experiments imply that: (1) a unique training sample will not be representative enough for certain areas, where the model will be unable to generate tree predictions, and (2) even for the areas for which it may be relatively representative it may not be representative enough to achieve predictions that are comparable to independent models. Therefore, these findings highlight the importance of considering the heterogeneity of areas when raining models for tree canopy prediction.

4. Application

This section describes an example application of the algorithm to estimate a panel of tree canopy coverage. The main objective is to compare the output generated by the algorithm with the panel acquired from the tree censuses. The comparison aims to demonstrate the reliability of the algorithm, highlight its potential uses, provide guidance on its limitations and offer insights into how to use the algorithm for effective and reliable urban canopy estimation. The city of New York is a unique area for experimentation due to its tradition of conducting street tree inventories every ten years since 1995, which allows for obtaining a comparison panel constructed from street inventory data. Hence, the application presented in this section compares this census data to the obtained by applying the workflow to NAIP imagery from New York City in 2006 and 2015. However, it is worth noting that while predicted data capture trees throughout the entire city similarly to Ventura et al. (2022), the census panel is limited to street trees only. Hence, to increase the comparability of both sets of results, the analysis removes all areas covered by parks owned by the City of New York.

The resulting estimates from the algorithm are at the pixel level, which has the advantage of reflecting the area covered by trees but is not directly comparable to tree counts. This section implements a simple approach to approximate tree counts from pixel data to provide additional comparisons. The conversion starts by vectorizing the tree pixel data, which generates a single square polygon for each tree pixel. The polygon data represents the same as the pixel data in a different format, whose advantage is that it allows combining all pixels into a single geometry object. This step reduces the complexity of the vector data by converting all pixels to a unique multi-geometry polygon. The final step explodes the resulting multi-geometry, which separates the polygon into individual geometries depending on adjacency. The last step involves exploding the multi-geometry object, which dissects the polygon into as many distinct geometries as different parts the polygon has. This explosion only separates non-touching geometries, which implies that some polygons that represent individual trees but are adjacent

may remain combined. Therefore, the vectorized tree counts area a lower bound to the number of trees estimated with the algorithm.

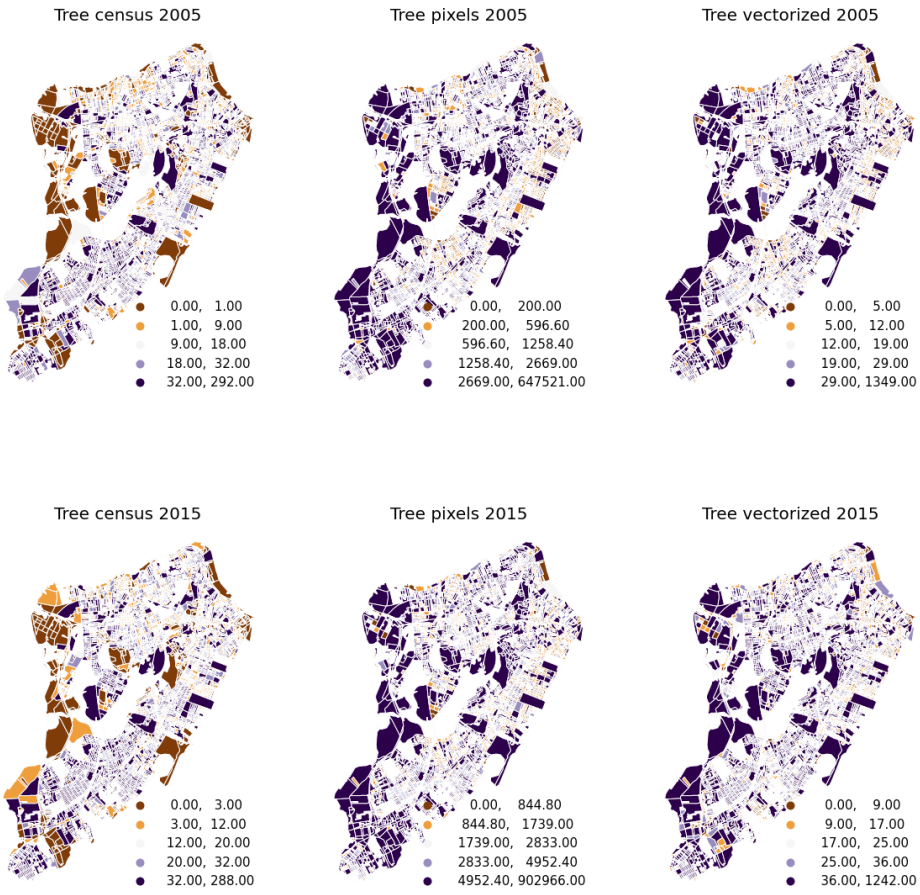
After constructing the canopy data, the next step in the analysis is to aggregate the tree canopy coverage estimates into consistent geographic areas that allow for comparisons within the city and across time. As the unit of intervention of public policy and urban planning are neighborhoods, the application starts by aggregating the data to the Census block 2010 definition. Census Blocks are the finest geographic subdivision of the Census, containing between 200-500 housing units, and are bound by geographic or jurisdictional elements. In the case of the Five Boroughs of New York, the average area of the blocks is 0.3 square kilometers. One advantage of aggregating data to Census Blocks or other geographic Census subdivisions, is the availability of Census information on demographic and socioeconomic characteristics, a requisite for designing policies that prioritize equity and address environmental justice concerns within the city[4].

In practice, each block each year is characterized by three canopy levels: the count of tree census and tree pixels that fall within the block, and the amount of vectorized trees that intersect the block. Figure 10 displays the tree canopy statistics in Staten Island for 2005 and 2015, which was selected as it highlights the discrepancies present even after eliminating New York City parks areas: areas with no street trees according to the Census are actually areas with high levels of canopy coverage in other types of green spaces not belonging to the City. In fact, while with the Census the estimated total number of trees in Staten Island's blocks was 97,506 in 2005, it raised to 115,236 using the vectorized tree count. The discrepancies imply that traditional tree census can undercount the number of trees and the growth rates, which in this case would be 3% according to the Census but 16% with the detected vectorized ones.

Table 5.1.4 contains descriptive statistics for the 2005 and 2015 canopy levels in each Borough. The table shows that the distribution of vectorized estimates is similar in terms of mean and median with the Census counts. However, the algorithm consistently detects higher maximum levels of trees for both years and exhibits higher standard deviations. The implication is that while the distribution of street trees is less varied, the canopy coverage estimated with the algorithm, which includes green spaces and privately owned trees, exhibits higher variability and maximum levels. The discrepancy highlights the importance of considering both public street trees, privately owned trees, and trees in green spaces to promote equality in urban tree canopy coverage. The table also indicates that the algorithm, both with vectorized counts and pixels counts, successfully captures the broad pattern of change between 2005 and 2015, which, according to the street census, is an increase in foliage in the city streets. Appendix Figure 5.1.2 provides additional evidence by plotting the difference between the 2015 count of vectorized

[4]An alternative approach would be to aggregate over a constant grid which preserves variation within Census units and exploits the high resolution of the aerial imagery. However, using any geographic unit can lead to some limitations, related to size and population variation within Census units or the choice of cell size, as prediction errors will magnify in smaller areas and larger units will erase the variation. The choice will depend on the research question at hand.

Figure 10: Block level tree canopy data



Notes: Figure shows the tree canopy levels at the Block level in Staten Island. First row represents that three measures (Street Census data, Pixel counts, Vectorized tree counts) for 2005 and the second row, for 2015.

Table 5: Areas with no trees

% 0 Trees	Blocks	Block groups	Tracts
2015 Census	0.15	0.01	0.01
2015 Vectorized	0.14	0.02	0.007

Notes: Table shows the percentage of units that have zero trees in 2015 according to street census and the vectorization, and separately for the different aggregation units.

trees and the Census. It shows that while differences tend to be small between the two counts, the vectorized counts are higher in large block groups, which typically contain large green spaces.

The next step of the analysis explores the sensitivity of the estimates to the area of the geographic units of aggregation. This issue is relevant as estimation errors will magnify in smaller units: labeling pixels incorrectly will lead to significant underestimations of coverage, which will be larger the smaller the area. The approach to tackle this question compares the estimates obtained through aggregation at the block level to those obtained at the block group and census tract levels, corresponding to immediately larger areas than blocks within the Census hierarchy. The comparison begins by comparing the number of neighborhoods with zero trees, a significant concern when aggregating data in small units. Moreover, blocks with no trees are particularly relevant for applications using growth rates, as such scenarios result in undefined rates for neighborhoods with no trees in the initial period, introducing bias in the estimation. Table 5 contains the percentage of units with no estimated trees according to the 2015 Census and vectorization with the three different geographic areas. As expected, the larger the aggregation areas, the lower the share of units with zero trees, according to the census and the vectorization. Also, the percentages obtained with both data sources are very similar. The similarities between the block group and tract estimates suggest those units have appropriate sizes to reduce the biases from zero estimated trees. Finally, the correlation between the three different canopy measures with each aggregation level is shown in 11. The results show that correlations increase with the unit area. Particularly, vector and census counts have a correlation of 0.8 in 2015, while this was 0.4 in blocks. The potential explanation for these higher correlations is that in larger units, the errors produced by the algorithm are relatively less important because the larger size of the units provides more opportunities to capture a higher number of trees, which can help mitigate the effects of errors. As previously discussed, tree pixels do not have a one-to-one mapping to Census counts, which explains the low correlations between both measures with all the levels. All in all, the findings suggest that block groups and tracts areas may be the optimal aggregation units, which would guide the grid construction to have areas ranging between 0.06-0.1 sq. km, corresponding to the median size of block groups and tracts.

Figure 11: Correlation matrices

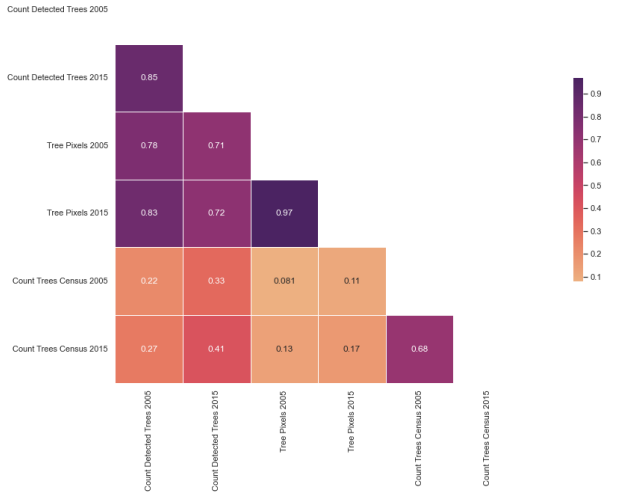


Figure A: Blocks

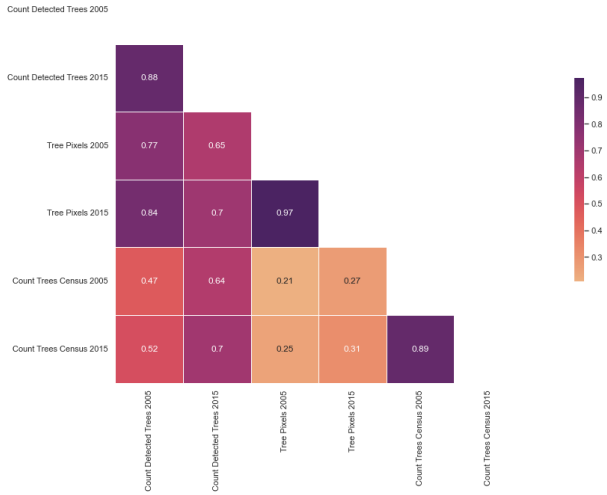


Figure B: Block groups

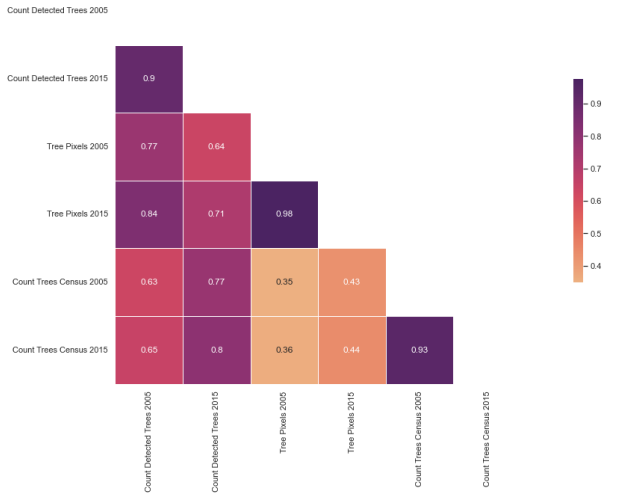


Figure C: Tracts

Notes: Figure shows correlation measures for the tree measures of tree canopy, with the three definition of aggregation areas.

5. Conclusion

Green spaces mitigate the adverse climatic and environmental effects of urban life and enhance the well-being of city dwellers. Understanding changes in tree coverage is also essential to design policies that guarantee equality in access to green amenities and ensure sustainable urban development. However, obtaining data on urban tree coverage over time is challenging, as traditional methods such as street tree census are time-consuming, costly, and cannot provide historical data. While recent machine-learning techniques offer promise, they require extensive expertise and modern data. This paper presents an innovative approach to rapidly and accurately detect tree data from aerial imagery for multiple time periods and in the presence of limited training data, addressing the need for cities to obtain tree coverage data efficiently.

This paper presents a novel, fully automated workflow for generating tree canopy panels. The proposed method builds upon the tree detection algorithm developed by Yang et al. (2009) and Bosch (2020) (YWPM&B), which has the notable advantage of only requiring natural-colored aerial images. To create ground-truth masks, the workflow leverages the reflectance of alive vegetation on near-infrared (NIR) light and employs histogram thresholding techniques to label pixels as trees or not trees. Although this multi-spectral data is only available in high-resolution imagery for recent periods, this paper demonstrates that it can be used to train models for earlier periods if the colors of images in both periods have been equalized. This paper also describes the implementation of the algorithm using images from the National Agricultural Imagery Product (NAIP) in two time periods for a set of urban areas in the United States, representing 7% of its urban area.

To assess the relevance of using the proposed workflow, a series of experiments are performed using urban areas that represent the different climatic zones of the U.S. The first set of experiments focuses on the creation of the ground-truth mask. One of the considered scenarios simplifies the creation of the masks by eliminating the image pre-processing steps. The results show that the simplification is less robust to data variation and more likely to introduce a higher number of false negatives. Also, training the YWPM&B with this data and comparing it to the ground-truth data obtained with the full pre-processing, shows that the simplification is associated with more false negatives. The last two experiments show that by not equalizing colors and using universal models rather than city models the model fails to predict tree presence in certain areas while doing worse than the complete workflow in other. Overall, the experiments results demonstrate the effectiveness and robustness of the proposed workflow in accurately detecting tree presence.

The final section of the paper highlights the advantages of the methodology by comparing it with the traditional street census approach. Specifically, it compares the estimated tree canopy coverage obtained using the proposed approach with the data from the 2005 and 2015 tree censuses conducted by the City of New York. The results demonstrate that the algorithm's predictions are generally consistent with the census data. However, the study also reveals the importance of including green coverage in other areas, such as private properties or other green

spaces, to accurately capture the urban tree canopy's evolution and changes, which cannot be achieved with census data. The section also discusses the transformation of pixel-level data into tree counts and the choice of the aggregating unit, highlighting that aggregating tree coverage data to the Census block-group and tract-level data reduces estimation biases.

In summary, this paper presents a novel methodology for constructing panels of tree canopy that have been demonstrated to be robust in estimation with high levels of accuracy, recall, and precision. The proposed approach captures broad changes in tree canopy similar to street census data but has the added advantage of providing a more comprehensive picture of urban tree canopy. Additionally, this paper shows how to work with the generated data by transforming pixel-level data to tree counts and selecting the appropriate aggregating unit. Future work will implement this methodology across all Metropolitan Statistical Areas of the United States and make the panel of tree canopy data available at block, block-group, and tract levels, as well as generate tree coverage grids for the areas. This research, while being methodological, has the potential of addressing multiple research questions related to the historical change in green amenities within and across cities, sustainability of urban development and environmental justice.

References

Karine RM Adeline, M Chen, X Briottet, SK Pang, and N Paparoditis. Shadow detection in very high spatial resolution aerial images: A comparative study. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80:21–38, 2013.

Sara Beery, Guanhang Wu, Trevor Edwards, Filip Pavetic, Bo Majewski, Shreyasee Mukherjee, Stanley Chan, John Morgan, Vivek Rathod, and Jonathan Huang. The auto arborist dataset: A large-scale benchmark for multiview urban forest monitoring under domain shift. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21294–21307, 2022.

Martí Bosch. Detectree: Tree detection from aerial imagery in python. *Journal of Open Source Software*, 5(50):2172, 2020. doi: 10.21105/joss.02172.

Guang Chen and Yi Shang. Transformer for tree counting in aerial images. *Remote Sensing*, 14(3), 2022. ISSN 2072-4292. doi: 10.3390/rs14030476. URL https://www.mdpi.com/2072-4292/14/3/476.

R.C. Gonzalez and R.E. Woods. *Digital Image Processing*. Pearson, 2018. ISBN 9780133356724. URL https://books.google.es/books?id=0F05vgAACAAJ.

Mark Grundland and Neil A Dodgson. Color histogram specification by histogram warping. In *Color Imaging X: Processing, Hardcopy, and Applications*, volume 5667, pages 610–621. SPIE, 2005.

S N H Syed Hanapi, S A A Shukor, and J Johari. A review on remote sensing-based method for tree detection and delineation. *IOP Conference Series: Materials Science and Engineering*, 705 (1):012024, nov 2019. doi: 10.1088/1757-899X/705/1/012024. URL https://dx.doi.org/10.1088/1757-899X/705/1/012024.

Benjamin A. Jones. Planting urban trees to improve quality of life? the life satisfaction impacts of urban afforestation. *Forest Policy and Economics*, 125:102408, 2021. ISSN 1389-9341. doi: https://doi.org/10.1016/j.forpol.2021.102408. URL https://www.sciencedirect.com/science/article/pii/S1389934121000149.

Arnon Karnieli, Nurit Agam, Rachel T. Pinker, Martha Anderson, Marc L. Imhoff, Garik G. Gutman, Natalya Panov, and Alexander Goldberg. Use of ndvi and land surface temperature for drought assessment: Merits and limitations. *Journal of Climate*, 23(3):618 – 633, 2010. doi: 10.1175/2009JCLI2900.1. URL https://journals.ametsoc.org/view/journals/clim/23/3/2009jcli2900.1.xml.

S. J. Livesley, E. G. McPherson, and C. Calfapietra. The urban forest and ecosystem services: Impacts on urban water, heat, and pollution cycles at the tree, street, and city scale. *Journal of Environmental Quality*, 45(1):119–124, 2016. doi: https://doi.org/10.2134/jeq2015.11.0567. URL https://acsess.onlinelibrary.wiley.com/doi/abs/10.2134/jeq2015.11.0567.

E. Gregory McPherson, Natalie van Doorn, and John de Goede. Structure, function and value of street trees in california, USA. *Urban Forestry &amp Urban Greening*, 17:104–115, jun 2016. doi: 10.1016/j.ufug.2016.03.013. URL https://doi.org/10.1016%2Fj.ufug.2016.03.013.

Dominic J. Morales. The contribution of trees to residential property value. *Journal of Arboriculture*, 6(11):305–308, 1980.

Mila Nikolova, You-Wei Wen, and Raymond Chan. Exact histogram specification for digital images using a variational approach. *Journal of Mathematical Imaging and Vision*, 46(3):309–325, 2013.

Kaori Otsu, Magda Pla, Andrea Duane, Adrián Cardil, and Lluís Brotons. Estimating the threshold of detection on tree crown defoliation using vegetation indices from uas multispectral imagery. *Drones*, 3(4), 2019. ISSN 2504-446X. URL https://www.mdpi.com/2504-446X/3/4/80.

Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979. doi: 10.1109/TSMC.1979.4310076.

Hans Pretzsch, Peter Biber, Enno Uhl, Jens Dahlhausen, Thomas Rötzer, Juan Caldentey, Takayoshi Koike, Tran van Con, Aurélia Chavanne, Thomas Seifert, Ben du Toit, Craig Farnden, and Stephan Pauleit. Crown size and growing space requirement of common tree species in urban centres, parks, and forests. *Urban Forestry Urban Greening*, 14(3): 466–479, 2015. ISSN 1618-8667. doi: https://doi.org/10.1016/j.ufug.2015.04.006. URL https://www.sciencedirect.com/science/article/pii/S1618866715000473.

Colleen E Reid, Jane E Clougherty, Jessie LC Shmool, and Laura D Kubzansky. Is all urban green space the same? a comparison of the health benefits of trees and grass in new york city. *International journal of environmental research and public health*, 14(11):1411, 2017.

Kirsten Schwarz, Michail Fragkias, Christopher G. Boone, Weiqi Zhou, Melissa McHale, J. Morgan Grove, Jarlath O'Neil-Dunne, Joseph P. McFadden, Geoffrey L. Buckley, Dan Childers, Laura Ogden, Stephanie Pincetl, Diane Pataki, Ali Whitmer, and Mary L. Cadenasso. Trees grow on money: Urban tree canopy cover and environmental justice. *PLOS ONE*, 10(4): 1–17, 04 2015. doi: 10.1371/journal.pone.0122051. URL https://doi.org/10.1371/journal.pone.0122051.

Dori Shapira, Shai Avidan, and Yacov Hel-Or. Multiple histogram matching. In *2013 IEEE international conference on image processing*, pages 2269–2273. IEEE, 2013.

Mardelle Shepley, Naomi Sachs, Hessam Sadatsafavi, Christine Fournier, and Kati Peditto. The impact of green space on violent crime in urban environments: an evidence synthesis. *International journal of environmental research and public health*, 16(24):5119, 2019.

Panu Srestasathiern and Preesan Rakwatin. Oil palm tree detection with high resolution multi-spectral satellite imagery. *Remote Sensing*, 6(10):9749–9774, 2014. ISSN 2072-4292. doi: 10.3390/rs6109749. URL https://www.mdpi.com/2072-4292/6/10/9749.

Kalyan Sunkavalli, Micah K Johnson, Wojciech Matusik, and Hanspeter Pfister. Multi-scale image harmonization. *ACM Transactions on Graphics (TOG)*, 29(4):1–10, 2010.

Jonathan Ventura, Milo Honsberger, Cameron Gonsalves, Julian Rice, Camille Pawlak, Natalie LR Love, Skyler Han, Viet Nguyen, Keilana Sugano, Jacqueline Doremus, et al. Individual tree detection in large-scale urban environments using high-resolution multispectral imagery. *arXiv preprint arXiv:2208.10607*, 2022.

Jan D. Wegner, Steve Branson, David Hall, Konrad Schindler, and Pietro Perona. Cataloging public objects using aerial and street-level images — urban trees. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6014–6023, 2016. doi: 10.1109/CVPR.2016.647.

Ben G. Weinstein, Sergio Marconi, Stephanie Bohlman, Alina Zare, and Ethan White. Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks. *Remote Sensing*, 11(11), 2019. ISSN 2072-4292. doi: 10.3390/rs11111309. URL https://www.mdpi.com/2072-4292/11/11/1309.

Ben G Weinstein, Sergio Marconi, Mélaine Aubry-Kientz, Gregoire Vincent, Henry Senyondo, and Ethan P White. Deepforest: A python package for rgb deep learning tree crown delineation. *Methods in Ecology and Evolution*, 11(12):1743–1751, 2020.

Ben G. Weinstein, Sarah J. Graves, Sergio Marconi, Aditya Singh, Alina Zare, Dylan Stewart, Stephanie A. Bohlman, and Ethan P. White. A benchmark dataset for canopy crown detection and delineation in co-registered airborne rgb, lidar and hyperspectral imagery from the national ecological observation network. *PLOS Computational Biology*, 17(7):1–18, 07 2021. doi: 10.1371/journal.pcbi.1009180. URL https://doi.org/10.1371/journal.pcbi.1009180.

Jinru Xue and Baofeng Su. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of sensors*, 2017, 2017.

Lin Yang, Xiaqing Wu, Emil Praun, and Xiaoxu Ma. Tree detection from aerial imagery. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '09, pages 131–137, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605586496. doi: 10.1145/1653771.1653792.

Appendix

5.1 Additional evidence and results

Figure 5.1.1: Aerial image for part of NYC

Notes: Figure shows a zoomed-in part of the aerial image for NYC in 2015.

Table 5.1.1: Performance - baseline model

City	Accuracy	Recall	Precision
Akron	0.85	0.89	0.72
Atlanta	0.83	0.76	0.80
Baltimore	0.92	0.85	0.82
Birmingham	0.80	0.80	0.63
Boston	0.94	0.88	0.80
Cambridge	0.94	0.82	0.77
Camden	0.92	0.82	0.79
Cleveland	0.84	0.79	0.68
Columbus	0.90	0.83	0.83
Dayton	0.88	0.59	0.80
Detroit	0.90	0.75	0.80
East St Louis	0.88	0.81	0.77
Flint	0.88	0.79	0.76
Kansas	0.87	0.75	0.76
Milwaukee	0.92	0.85	0.83
Nashville	0.93	0.87	0.83
New Haven	0.93	0.75	0.89
New Orleans	0.92	0.83	0.82
Oakland	0.92	0.45	0.74
Richmond	0.92	0.87	0.76
San Francisco	0.93	0.70	0.69
Seattle	0.93	0.70	0.78
St Louis	0.88	0.72	0.75
Toledo	0.90	0.86	0.80
Trenton	0.91	0.79	0.83
Bronx	0.93	0.81	0.75
Brooklyn	0.95	0.76	0.81
Buffalo	0.92	0.67	0.78
Manhattan	0.94	0.88	0.78
Queens	0.93	0.72	0.83
Rochester	0.87	0.77	0.60
Staten Island	0.93	0.84	0.76
Syracuse	0.90	0.67	0.75
Westchester	0.88	0.75	0.76
Chicago	0.92	0.67	0.85
Los Angeles	0.91	0.52	0.83
Somerville	0.96	0.92	0.89

Notes: Table displays the performance results of running the algorithm using a 5% of tiles.

Table 5.1.2: Performance, 10% sample

City	Accuracy	Recall	Precision
Akron	0.86	0.83	0.76
Birmingham	0.84	0.78	0.68
Bronx	0.92	0.86	0.68
Brooklyn	0.95	0.70	0.79
Flint	0.89	0.78	0.81
Manhattan	0.93	0.85	0.73
New Orleans	0.93	0.83	0.83
Queens	0.94	0.74	0.81
San Francisco	0.93	0.61	0.67
Seattle	0.90	0.56	0.81
Staten Island	0.92	0.85	0.72
Westchester	0.89	0.78	0.78

Notes: Table displays the performance results of running the algorithm using a 10% of tiles in the areas used for experimentation.

Table 5.1.3: Experimentation: modifying ground-truth masks 10% sample

	Exp. 1: NDVI Threshold			Exp. 2: NDVI threshold vs. full processing		
City	Accuracy	Recall	Precision	Accuracy	Recall	Precision
Akron	0.70	0.29	0.96	0.82	0.91	0.66
Birmingham	0.73	0.59	0.63	0.84	0.86	0.66
Bronx	0.93	0.87	0.85	0.91	0.94	0.68
Brooklyn	0.94	0.78	0.85	0.93	0.82	0.71
Flint	0.80	0.51	0.88	0.86	0.95	0.66
Manhattan	0.96	0.85	0.89	0.94	0.84	0.74
New Orleans	0.82	0.41	0.90	0.91	0.95	0.69
Queens	0.94	0.77	0.89	0.92	0.88	0.71
San Francisco	0.85	0.18	0.54	0.94	0.52	0.66
Seattle	0.75	0.25	0.84	0.85	0.72	0.56
Staten Island	0.94	0.82	0.88	0.93	0.93	0.71
Westchester	0.89	0.84	0.85	0.82	0.94	0.61

Notes: Table displays the performance results of running the algorithm using only double thresholding on NDVI (Exp. 1) and training the algorithm defining with only NDVI thresholding but comparing to ground-truth data using all the full pre-processing (Exp 2.) on a 10% of tiles.

Table 5.1.4: Descriptive statistics

	Borough	2005 Census	Vectorized 2005	Pixels 2005	2015 Census	Vectorized 2015	Pixels 2015
Min.	Bronx	0	0	0	0	0	0
	Brooklyn	0	0	0	0	0	0
	Manhattan	0	0	0	0	0	0
	Queens	0	0	0	0	0	0
	Staten Island	0	0	0	0	0	0
Mean	Bronx	11	7	1,187	15	11	2,072
	Brooklyn	14	11	949	17	18	1,904
	Manhattan	13	5	339	17	9	1,014
	Queens	16	13	1,409	16	18	2,273
	Staten Island	20	23	4,221	20	27	6,360
Median	Bronx	7	4	197	13	8	868
	Brooklyn	11	6	178	16	15	983
	Manhattan	8	2	32	15	6	428
	Queens	13	10	527	15	15	1,378
	Staten Island	14	15	876	16	21	2,239
Maximum	Bronx	162	1,054	683,839	287	837	811,663
	Brooklyn	330	2,911	1,017,809	201	1,841	1,275,645
	Manhattan	202	211	60,442	150	232	65,045
	Queens	196	1,570	408,700	175	1,340	563,189
	Staten Island	292	1,349	647,521	288	1,242	902,966
Std. Dev.	Bronx	13	18	10,000	14	17	12,007
	Brooklyn	15	37	15,248	14	34	16,796
	Manhattan	15	9	1,831	15	11	2,355
	Queens	16	24	8,483	14	23	8,967
	Staten Island	24	46	23,805	21	37	28,885

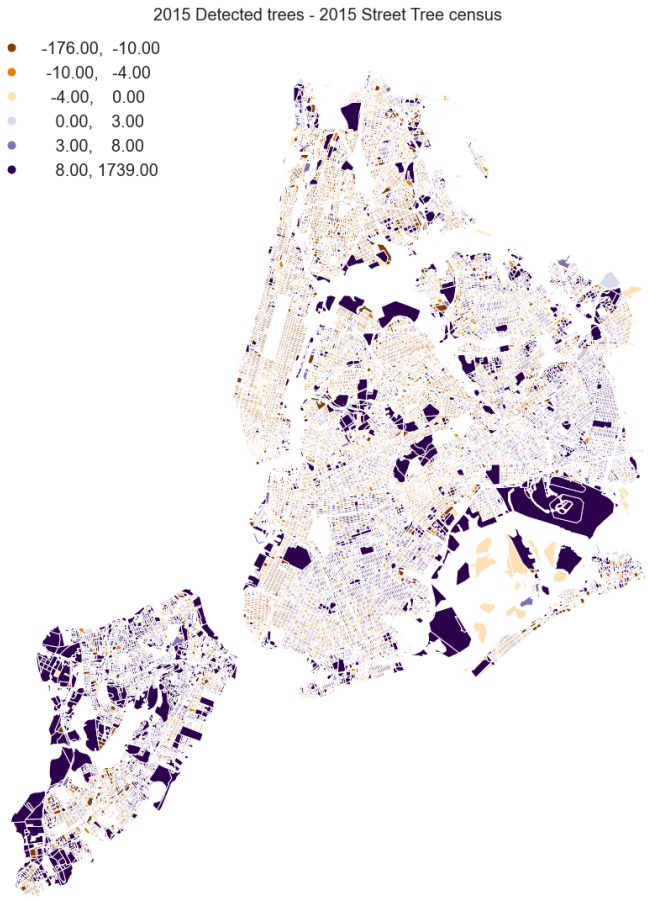
Notes: Table displays descriptive statistics of tree coverage, aggregated at the block level, in the city of New York.

Table 5.1.5: Descriptive statistics for area (sq.km) of blocks in NYC

	No tree pixels in 2005	Rest of blocks
Count	2,102	26,535
Mean	0.0119	0.0200
Std.Dev.	0.0079	0.0443
Minimum	0.0004	0.0005
25%	0.0071	0.0115
50%	0.0113	0.0155
75%	0.0155	0.0196
Maximum	0.1694	1.9966

Notes: Table displays descriptive statistics for the area (in sq.km) of blocks in New York City, separately for areas without any tree pixel in 2005 and the rest.

Figure 5.1.2: Difference in vectorized counts and Census counts, 2015



Notes: Figure shows the difference in the block counts between vectorized trees and Census trees in 2015.