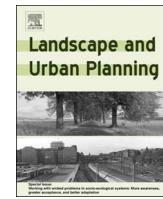




Senseable City Lab :::: Massachusetts Institute of Technology

This paper might be a pre-copy-editing or a post-print author-produced .pdf of an article accepted for publication. For the definitive publisher-authenticated version, please refer directly to publishing house's archive system



Research Paper

Green streets – Quantifying and mapping urban trees with street-level imagery and computer vision

Ian Seiferling^{a,b,*}, Nikhil Naik^c, Carlo Ratti^a, Raphaël Proulx^b^a Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Room 10-485, 77 Massachusetts Avenue, Cambridge, MA, 02139, United States^b Canada Research Chair in Ecological Integrity, Centre de Recherche sur les Interactions Bassins Versants-Ecosystèmes Aquatiques, Université du Québec à Trois-Rivières, 3351 Boulevard des Forges, Trois-Rivières, Québec, G9A 5H7, Canada^c MIT Media Lab, 75 Amherst St, Cambridge, MA 02139, United States

ARTICLE INFO

Keywords:
 Urban trees
 Computer vision
 Streetscapes
 Tree cover
 Greenspace

ABSTRACT

Traditional tools to map the distribution of urban green space have been hindered by either high cost and labour inputs or poor spatial resolution given the complex spatial structure of urban landscapes. What's more, those tools do not observe the urban landscape from a perspective in which citizens experience a city. We test a novel application of computer vision to quantify urban tree cover at the street-level. We do so by utilizing the open-source image data of city streetscapes that is now abundant (Google Street View). We show that a multi-step computer vision algorithm segments and quantifies the percent of tree cover in streetscape images to a high degree of precision. By then modelling the relationship between neighbouring images along city street segments, we are able to extend this image representation and estimate the amount of perceived tree cover in city streetscapes to a relatively high level of accuracy for an entire city. Though not a replacement for high resolution remote sensing (e.g., aerial LiDAR) or intensive field surveys, the method provides a new multi-feature metric of urban tree cover that quantifies tree presence and distribution from the same viewpoint in which citizens experience and see the urban landscape.

1. Introduction

With the growing consensus that nature and multi-functional ecosystems are intrinsic to sustainable cities, decision makers, designers and the broader public alike are looking to trees as urban keystone flora that provide natural infrastructure and services – to reduce air pollution, support biodiversity, mitigate heat island effects, increase land value, improve aesthetics and even improve human health (Kardan et al., 2015; Lothian, 1999; Lovasi et al., 2008; McPherson et al., 1997; Nowak, Hirabayashi, Bodine, & Greenfield, 2014; Thayer and Atwood, 1978). Urban tree effects may even extend to cultural and psychological behaviours with, for example, a high abundance of street trees being linked to urban scenes that were perceived to be safe (Naik, Philipoom, Raskar, & Hidalgo, 2014). The fact remains however that urban trees come with costs and are currently threatened by climate change, pests and diseases. Conflicting land uses and cost-benefit tradeoffs cause contention at many levels of society. Such contentions can be alleviated through a better understanding of the role of trees in the complex and cluttered landscapes that are cities. To this end, tools to quantify and

monitor presence, abundance and health of urban trees are needed. Governments, particularly cash-strapped ones, are evermore looking for low-cost ways to establish baseline data, manage and engage the public on urban trees.

Traditionally, urban tree cover has been quantified using coarse-scale methods developed for naturally forested landscapes and exposure to “nature” as an urban quality indicator has been quantified by measuring the total land area covered by greenspace (*i.e.*, city park area) in cities (Fuller and Gaston, 2009; Richardson, Pearce, Mitchell, & Kingham, 2013; Schroeder, 1986). In either case, these methods primarily rely on long-range remotely-sensed image processing to classify landcover (*i.e.*, satellite imagery such as LANDSAT, ortho-aerial photographs or, more recently, LiDAR) (Homer et al., 2007) or data derived from field surveys (Kardan et al., 2015). Substantial drawbacks exist within each case, many of which present particular challenges in an urban context. For example, traditional remote-sensing techniques for vegetation cover have, most often, been based on moderate-resolution imagery (e.g., 30 m in the case of openly available data) which has limited utility at the scale of cities. Recent

* Corresponding author at: Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Room 10-485, 77 Massachusetts Avenue, Cambridge, MA, 02139, United States.

E-mail address: ianseifs@mit.edu (I. Seiferling).

efforts exploiting high resolution active sensing like LiDAR are proving well-suited for urbanscapes (MacFaden, O'Neil-Dunne, Royar, Lu, & Rundle, 2012), however they can be hindered by specialized proprietary software, high data-acquisition costs and significant labour inputs. On the other hand, field-based surveys lack the automation and the scale of big data sets (*i.e.*, low-throughput), are prone to sampling errors (Dickinson, Zuckerberg, & Bonter, 2010) and require enormous organizational efforts. These methodological impediments also make it difficult to achieve periodic resampling to assess changes in tree cover and health over time.

Chiefly through machine learning models, computer vision scientists are teaching computers to see the world at astounding rates of success. However, few disciplines outside of the strict artificial intelligence fields (*e.g.*, robotics, driverless cars, software) have utilized these advancements. One of the few examples bridging ecology and computer vision technologies is the mobile app, *Leafsnap*, which identifies plant species using automatic visual recognition (Kumar et al., 2012). If a computer can learn to detect and quantify features of an environmental scene from digital photographs (*i.e.*, scene understanding), it stands that those algorithms can be used to objectively quantify real-world features and their spatial distribution within a landscape for a multitude of applications. For instance, Naik et al. have developed computer vision algorithms that process street-level imagery to quantify urban appearance (Naik et al., 2014), urban change (Naik, Kominers, Raskar, Glaeser, & Hidalgo, 2015), or even socio-economic indicators (Glaeser, Kominers, Luca, & Naik, 2015). Opportunely, we now also have access to entire cities in the form of geo-tagged, street-level images.

Using Google Street View images that represent a ground-based perspective of city streets – *streetscapes* – and which cover a city-wide extent, we develop and test a new method of rapid quantification and mapping of urban vegetation, specifically trees. The method applies a trained predictor to segment the amount of tree cover in a given image of a city streetscape using multiple image features. We aim to demonstrate that we can quantify the presence and perceived cover of street-side trees with high spatial resolution at the city-scale by: *i*) sampling a series of sequential neighbouring image scenes of the streetscape; *ii*) predicting the amount of tree cover present in them and; *iii*) modelling the relationship between the tree cover of these neighbouring view-points. To estimate the accuracy and utility of this approach we compare our method to contemporary remote-sensing techniques used to estimate urban tree canopy cover (*i.e.*, object-based image analysis (OBIA) of high-resolution LIDAR data and multispectral imagery).

The goal of this study is to present a novel method of measuring trees in a city at extremely high-throughput; one that may not replace existing techniques, but offers clear benefits such as being relevant to the human perspective (*i.e.*, *the perceived tree cover*), cheap, independent of proprietary software and easily scaleable across cities.

2. Methods

2.1. Study areas and image datasets

We collected data on urban tree cover by using 456,175 geo-tagged images from the two cities of New York (336,998 images) and Boston (119,177 images) in the United States. However, for the vast majority of the results presented, we focus on New York because the best-suited tree canopy cover maps and street tree survey data we could acquire were of New York. Images were sourced from the Google Street View (GSV) application program interface (API) (Google Inc., 2014), were acquired in 2014 and represent a ground-level, side-view perspective of the city streetscape (Fig. 1C). All image collection points along city roads were downloaded for a target city and this resulted in a GSV image roughly every 15 m along a given roadway; these image samples are hereafter referred to as *GSV sampling points*. However, due to the protocol of the GSV system the 15 m interval could deviate by

approximately +/– 5 m. Given this, we define a neighbour sample points as two GSV sampling points on the same road segment and a minimum of 10 m and maximum of 20 m apart. Some GSV sampling points, road segments or areas of the city did not have data for various reasons (*e.g.*, corrupt or missing data, no-coverage area). Notwithstanding those instances, the sampling regime covered the full extent of the cities' official boundaries, though for the case of New York it did not cover Staten Island (Fig. 1A & B).

Each digital photograph (Red-Green-Blue color channel jpeg image) was acquired from the GSV API at a resolution of 400 by 300 pixels, at a 90° horizontal field of view, 90° heading (east) and a 10° pitch. The level of pitch was chosen in order to optimize the capture of the streetscape (*i.e.*, decrease the amount of foreground composed only of roadway). Fixing the image heading to 90° east for every sampling point allowed us to compare how the road-to-image orientation would affect the metrics and, ultimately, the ability to estimate tree cover. As such, all sampling points were grouped into one of four categories based on their road orientation, given 22.5° intervals around 360°: 1) N-S: GSV sampling points lying on roads that are oriented in a north-south direction ($\pm 22.5^\circ$ from 0° or 180°); 2) E-W: GSV sampling points lying on roads oriented in an east-west direction ($\pm 22.5^\circ$ from 90° or 270°); 3) NW-SE: GSV sampling points lying on roads oriented in a diagonal northwest-southeast direction ($\pm 22.5^\circ$ from 135° or 315°); 4) NE-SW: GSV sampling points lying on roads oriented in a diagonal northeast-southwest direction ($\pm 22.5^\circ$ from 45° or 225°).

In order to estimate the real-world surface area covered by each GSV image, we modelled the 2-dimensional (horizontal and vertical) surficial field of view (FOV) represented in an image at each sampling point; *i.e.*, the camera's horizontal field of view (90°) and depth of field projected onto the earth's surface. We computed this FOV polygon for each GSV sampling point which was then projected on the horizontal surface plane to associate a surface area with the sampling point (Fig. 1d). The length of the polygon (*i.e.*, length of the right-angle bisector) represents the approximate image depth of field. However, in reality the depth of field varies with the presence, size and proximity of objects occluding the horizon. We assume that a given length should, on average, be representative of an urban streetscape. Therefore, we varied this depth of field parameter and created four levels: 15 m, 25 m, 35 m and 45 m from the GSV sampling point. In addition to the road-to-camera orientation groups, we run our analysis at each of these depth of field levels in order to determine which provides the best spatial context for predicting real-world tree cover.

2.2. Tree detection using computer vision

We estimated the total area covered by trees in each image by applying a multi-step image segmentation method developed by Hoiem et al. (Hoiem, Efros, & Hebert 2005). On a per-image basis, the objective of the method is to model geometric classes that depend on the orientation of a physical object with relation to the scene and with respect to the camera. Specifically, each image pixel is classified into one of a few geometric classes: *i*) the ground plane; *ii*) surfaces that stick up from the ground (vertical surfaces); *iii*) part of the sky plane. Further, vertical surfaces are subdivided into planar surfaces facing left, right or towards the camera and either porous (*e.g.* trees and their leafy vegetation) or solid (*e.g.* a person or lamp post) non-planar surfaces. Although this recognition approach differs from those that instead model semantic classes (*e.g.*, car, house, person, vegetation), it has proven exceptionally powerful and efficient in cluttered outdoor scenes like urban streetscapes and, most relevant to our application here, in distinguishing human built structures from natural ones like trees.

The algorithm operates by first grouping image pixels into *superpixels*, which are groups of pixels assumed to share a single label (*e.g.*, ground or sky) and respect coarse-level segment boundaries (*e.g.*, edges) (Felzenszwalb & Huttenlocher, 2004). The algorithm then groups regions of the image into homogenous segments using a

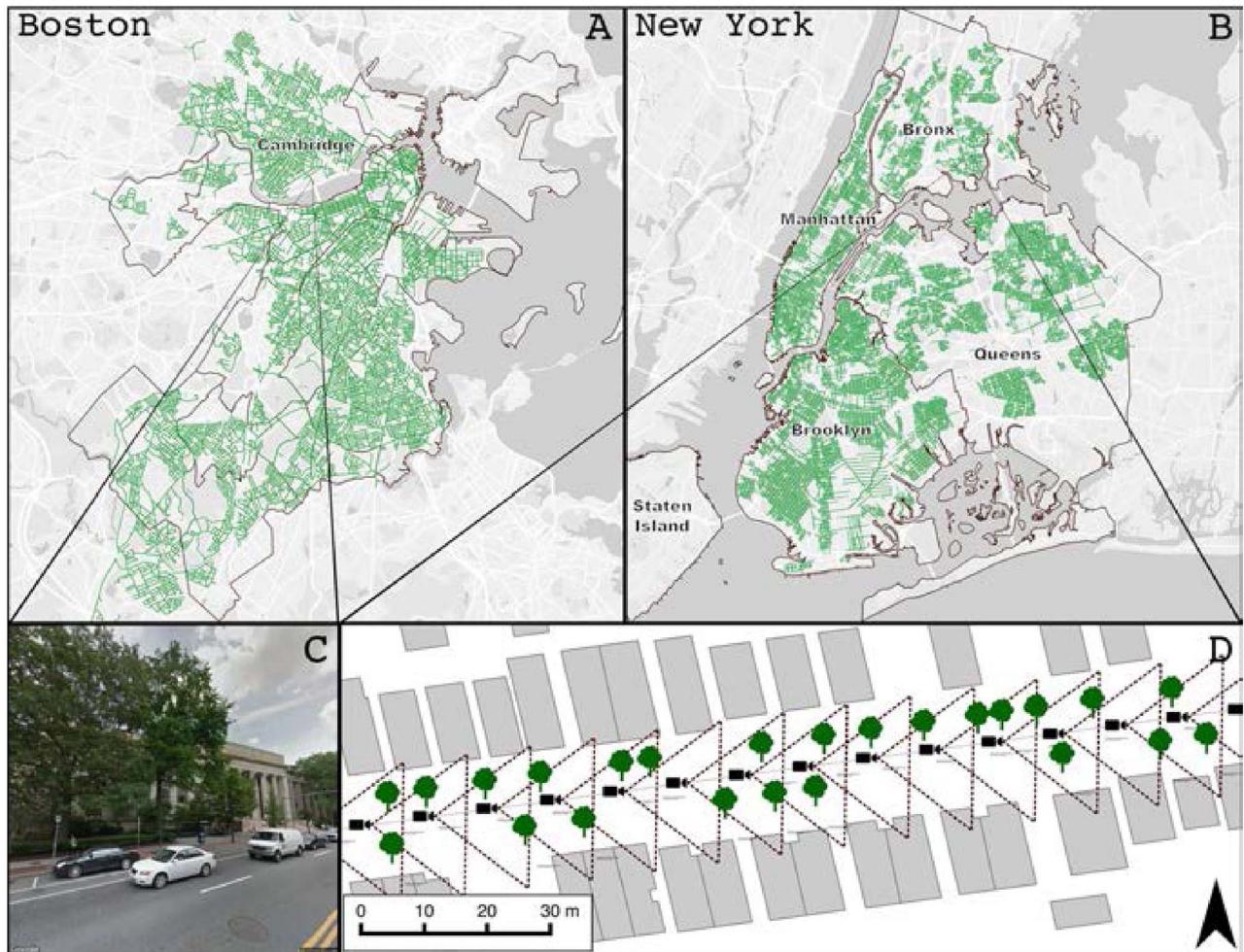


Fig. 1. Map and examples of the GSV image sampling extent, distribution, images and sampling design. **A.** a map of the city of Boston showing the extent and distribution of the GSV sampling points (green circles). **B.** a map of the city of New York showing the extent and distribution of GSV sampling points (green circles). **C.** an example of one GSV sample image from Boston representing a streetscape scene given the image orientation parameters. **D** an example of a street segment (in this case, an east-west orientation) in New York city illustrating the GSV sampling point design wherein a sequence of neighbouring sampling points (black camera icons) are located approximately 15 m apart along the street and each have an associated field of view polygon (dashed lines; here showing just the 15 m FOV-level) with a 90° heading. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

standard segmentation algorithm, but generates multiple hypothesis or combinations of these rough segmentations as it remains unknown which have been labelled correctly. Thus, a set of image features are computed at both the level of the super-pixel and the larger region segments. Using training image data of ground-truthed, labelled urban scenes, learning the parameters to predict final labels operates at two stages: *i*) Grouping super-pixels is learned by estimating the likelihood that two super-pixels belong in the same region based on their features. The multiple segmentation hypotheses are then generated by varying the number of regions and the initialization of the algorithm. *ii*) The final labelling of the geometric classes of image segments is learned by computing the features for each region and labelling them with a geometric class based on likelihood functions (*i.e.*, the likelihood that super-pixels have the same label and the confidence in each label). Once labelled in this fashion, the optimal likelihood functions are then learned through training (SI, Appendix).

With our images segmented and geometric classes labelled by this procedure, we applied the semantic labels to each pixel accordingly: ground, sky, building and trees. The percent of tree cover, ground, sky and building in an image were calculated as the total number of pixels belonging to that class divided by the total number of image pixels.

Since this method is well established in the computer vision field, we did not retain the pixel-level classification results. Due to this

approach we were not able to visually inspect the image segmentations that the algorithm performed and report a pixel-wise misclassification statistic. We refer to Hoiem et al. (2005) and Naik et al. (2015) for benchmarked classification statistics of the algorithm. Nonetheless, in order to validate the robustness of this approach we took a random sample of 100 Boston city GSV images and derived two estimates of the percent of tree cover in an image by pixel masking to compare to the output of the learning algorithm: *i*) we performed manual pixel masking by tracing all tree components in an image and summed the number of pixels falling inside and; *ii*) we computed a single-feature binary excess green index (Meyer & Neto, 2008; Proulx, Roca, Cuadra, Seiferling, & Wirth, 2014) which derives the proportion of green within an RGB image. Concerning the latter, similar single-feature methods have been used before to estimate vegetation presence in digital photographs (Li et al., 2015; Proulx et al., 2014; Yang, Zhao, McBride, & Gong, 2009). Considering the former, we performed the manual pixel masking under two schemes: *i*) a conservative estimation where only tree species and only their clearly defined components were traced, while shrubs, small woody vegetation and distant, poorly defined trees were excluded and; *ii*) a liberal estimation where tree-like, small woody vegetation was also included if it was clearly visible and, likewise, distant trees were included if we judged them very likely to be trees.

2.3. Modelling streetscape tree cover

Given an algorithm that quantifies the amount of area covered by trees in a 2D photograph, the remaining challenge we address is to relate such a value to a measure of real-world tree cover. Specifically, we test if and how well our metric of streetscape tree cover may estimate true percent tree canopy cover of the same area by modelling the relationship between them.

We derive the dependent variable of true percent canopy cover at each GSV sampling point FOV from a high resolution (3 ft.) and comprehensive land-cover map for New York developed by MacFaden et al. (MacFaden et al., 2012). Derived from LiDAR data and multi-spectral imagery acquired in 2010, the map represents the most up to date and high resolution data on New York's tree canopy. We compute the percent canopy cover inside each i^{th} GSV sampling point FOV, at each j^{th} FOV-level:

$$\% \text{ tree canopy cover}_{i,j} = \left(\frac{\text{number of pixels classified as tree inside } FOV_{i,j}}{\text{total number of pixels inside } FOV_{i,j}} \right) \times 100 \quad (1)$$

Simply measuring the tree cover from a 2D image poses two challenges in terms of predicting a real-world representation of it. First, the image representation compresses three dimensions into two (horizontal and vertical) and so a simple sum of same-classed pixels does not provide complete information on object depth or volume. This also means that trees that are occluded by other trees or objects can not be measured. Considering that this method is limited to streetscapes, this later effect is limited since urban street trees are typically not crowded and their general arrangement is one layer (row) of trees backed by buildings. Second, the amount of area covered by trees in an image will be highly dependent on the proximity of the tree(s) to the camera. Thus, the image tree cover is somewhat independent of their real-world size; all else equal, foliage that is closer to the camera will cover more of the image area. Both issues are a problem of perspective and location with respect to the camera at a given GSV sampling point.

We address these problems of perspective with the hypothesis that additional information about a scene can be gained from overlapping, neighbouring images which differ in their perspective and proximity to the same object. The amount of overlap between neighbouring images will depend on the road-to-image orientation since our image directions were fixed at a 90° east heading, while the road orientation varied (see Fig. 1D). We do not perform any stereoscopic interpretation explicitly, but by adding information from neighbouring images (i.e., the image-derived percent tree cover values), the real-world representation of a given scene may be approximated by an aggregation of all neighbouring GSV images which capture a portion of that scene.

More formally, we assume a linear relationship between the scene information captured by a central GSV sampling point, hereafter termed the “node”, and the contributions of its neighbouring GSV sampling points. We consider neighbour GSV points to be GSV sampling points on the same road segment as the node point and that are adjacent to the node in either direction. We estimated the average range of view for a typical GSV scene to be roughly 45 – 60 m by manually inspecting neighbouring GSV images. Given this range, we considered two neighbour points to either side of the node (four neighbours per GSV node point in total). We acknowledge that this parameter could be varied and tested in further applications. We establish a relationship between the real-world street tree cover and the GSV images of it:

$$\beta_i = \sum \omega_{i,n} X_{i,n} \quad (2)$$

where, β_i is the real-world measure of tree cover for the given scene i , X is the percent tree cover in the GSV images for the node i and neighbouring n sampling points and ω is a weighting factor representing the relative contribution of each neighbour, X_i . Since the distance interval between GSV sampling points is not always constant, we may

also normalize the contribution of each neighbour by dividing each neighbour term by its distance (δ) to the node point. Thus, expanding the equation and given that we include a total of four neighbouring points, with $-n$ neighbouring points below the node point and $+n$ neighbouring points above the node point, i :

$$\beta_i = \frac{\omega_1 X_{i,-2}}{\delta_{i,-2}} + \frac{\omega_2 X_{i,-1}}{\delta_{i,-1}} + \omega_3 X_i + \frac{\omega_4 X_{i,1}}{\delta_{i,1}} + \frac{\omega_5 X_{i,2}}{\delta_{i,2}} \quad (3)$$

To determine the weighted contributions of each image in a node-neighbour series, we learn the weighting factors by solving the linear system of equations for ω :

$$\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} \frac{X_{i,-2}}{\delta_{i,-2}} & \frac{X_{i,-1}}{\delta_{i,-1}} & X_i & \frac{X_{i,1}}{\delta_{i,1}} & \frac{X_{i,2}}{\delta_{i,2}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{X_{i,-2}}{\delta_{i,-2}} & \frac{X_{i,-1}}{\delta_{i,-1}} & X_i & \frac{X_{i,1}}{\delta_{i,1}} & \frac{X_{i,2}}{\delta_{i,2}} \end{bmatrix} \times \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \end{bmatrix} \quad (4)$$

Once the weighting factors are learned, they may then be plugged into Eq. (3) and standardized to compute the neighbour-weighted percent tree cover score at each GSV sampling point – the *streetscape tree cover*.

Finally, we model the relationship between our streetscape tree cover metric and the true percent tree canopy cover using multiple linear, least squares regression and cross-validate using 4-fold cross validation. In the regression models we also include the other semantic streetscape classes (i.e., percent building, ground and sky) as additional predictor variables under the hypothesis that the relative proportions of these features in an urban scene can provide descriptive information on the spatial arrangement, location and, hence, size of trees. Here we subset the data (predicted streetscape tree cover values and associated neighbour-weighted percent tree cover score) into a training set to build and validate the linear regression model and then test the predictive power of this model on a test set of new data.

We also test this relationship at the city district-levels of dynamic block, community district, school district and borough. A dynamic block (also known as atomic polygon) is the smallest unit in the city geospatial data. We compute the total percent tree canopy cover per block unit at each district-level (i.e., percent canopy cover in a given district polygon from the vertical view) and the associated mean of streetscape tree cover for all GSV sampling points (i.e., street-level view) inside each block unit (polygon).

In order to minimize the multiple sources of systemic error when relating the streetscape tree cover metric from GSV images to the percent tree canopy cover, we perform a set of filtering steps prior to the regression analysis (SI, Appendix). All image analysis and processing, geospatial analyses and statistical modelling were performed in the R software environment (R Core Team, 2013), Matlab (MATLAB, 2015), C++ and Python programming language.

3. Results

3.1. Manual pixel comparison

The streetscape tree cover algorithm applied to single GSV images most closely correlated with our liberal scheme of manual pixel masking in the streetscape images (adjusted r-square = 0.98) (Fig. 2). Though still highly correlated, the simple, single-feature automated green mask method did not relate as well to our streetscape tree cover estimator (adjusted r-square = 0.79). The conservative manual tree masks also related very closely to our tree detector (adjusted r-square = 0.97).

3.2. Predicting urban tree canopy cover

The results showed that percent canopy cover derived from the landcover map is relatively low for any given location in New York City;

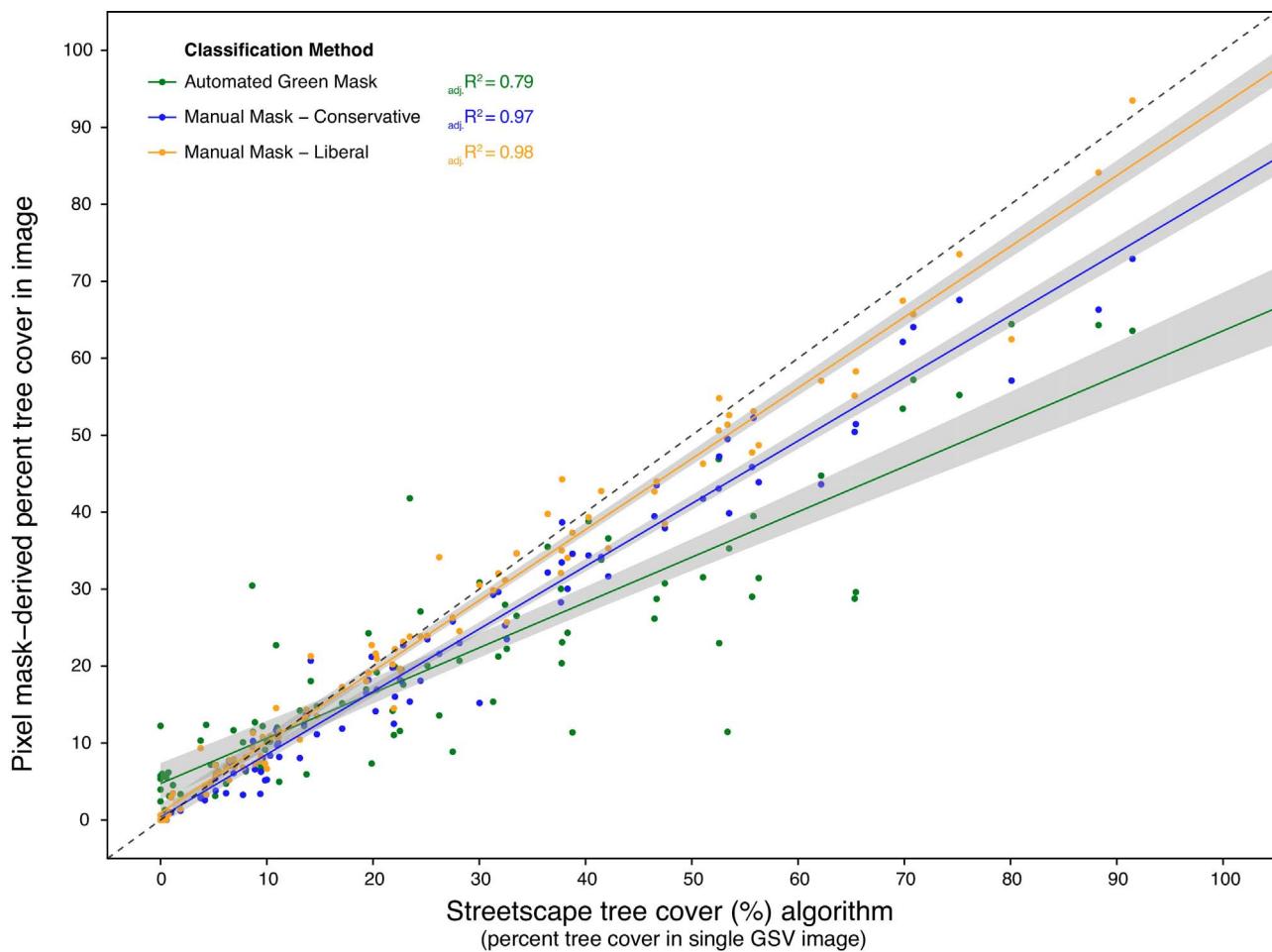


Fig. 2. Scatter plot with fitted linear regression lines for the relationships between the percent tree cover for a single GSV image as estimated by the streetscape tree cover algorithm (x-axis) and as estimated by the three pixel-masking methods: automated green mask (green circles and line), conservative manual mask (blue circles and line) and liberal manual mask (orange circles and line). The adjusted r -square values of the regressions are shown in their matching colors. The shaded grey represent the 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that is, the distribution of tree canopy cover was skewed to the left. For this reason, and to conform to the assumptions of homoscedasticity, we log-transformed the dependent variable of true canopy cover and used this semi-log structure in the final regression models. Moreover, there was a high number of data points with low and fractional percent tree canopy cover values (e.g., between 0 and 1 or 2 and 3% for example). These fractional differences are trivial, yet can amplify the importance of the data points when log-transformed. Therefore, we binned the response values by rounding them to the nearest integer.

Based on preliminary scatterplots and regression analysis, it was clear that the road-to-image orientation category 2 (E-W road orientation and parallel east camera heading) showed the strongest potential for predicting tree cover. We, henceforth, present and discuss results belonging to this category.

The relationships at all FOV-levels followed a nonlinear or curve-linear pattern, as would be predicted by the semi-log structure of the modelled relationship (i.e., the response variable of percent tree canopy cover is log-transformed and the predictor streetscape tree cover variable was untransformed). The best fitting regression models (*i.e.*, the highest r -squared and lowest root mean squared error (RMSE)) were at the 25 m and 35 m FOV-levels, with the 35 m level performing slightly better (Fig. 3 & Table 1). At the smaller FOV-levels (15 and 25 m) there was a high amount of false negative errors — sampling points with greater than zero streetscape tree cover predicted, but no trees (0% true canopy cover) in the associated FOV. Applying the neighbour-weighting procedure (Eqs. (2) & (3)) significantly improved

the predictive power of the streetscape tree cover estimations relative to using only an unweighted single node-GSV image-based value (Table 1).

Model residuals of the final regression models for the 35 m FOV had a mean close to 0 and followed a near-normal distribution, though the error variance was not perfectly constant. This was likely due partly to the high and skewed residual error at low, and in particular zero%, true canopy cover values where our predictor was detecting a range of tree cover values. The final regression model generalized well to a set of unseen test data (Fig. 3 & Table 1), having comparable r -square and root mean square error values (RMSE) between the training data ($r^2 = 0.74$, RMSE = 0.27) and the test data ($r^2 = 0.73$, RMSE = 0.28). When using the more rigorously filtered data subset that was used to learn the weighting factors, the model performance increased significantly ($r^2 = 0.81$, RMSE = 0.23).

Scaling the sampling units across district-levels resulted in an increasing predictive power of the method to estimate true canopy cover (Fig. 4 & Table S2). We note that for these results we enforced a sample size cutoff value for each district-level (see SI, Appendix for further details). This is why, for example, the Bronx borough is missing from that district-level at this time. Nonetheless, at the borough district-level, the results showed a very strong relationship between the mean streetscape tree cover and the total tree canopy cover for each borough (Fig. 4 & Table S2). It must be noted however, that there were only three boroughs remaining after the filtering steps (SI, Appendix).

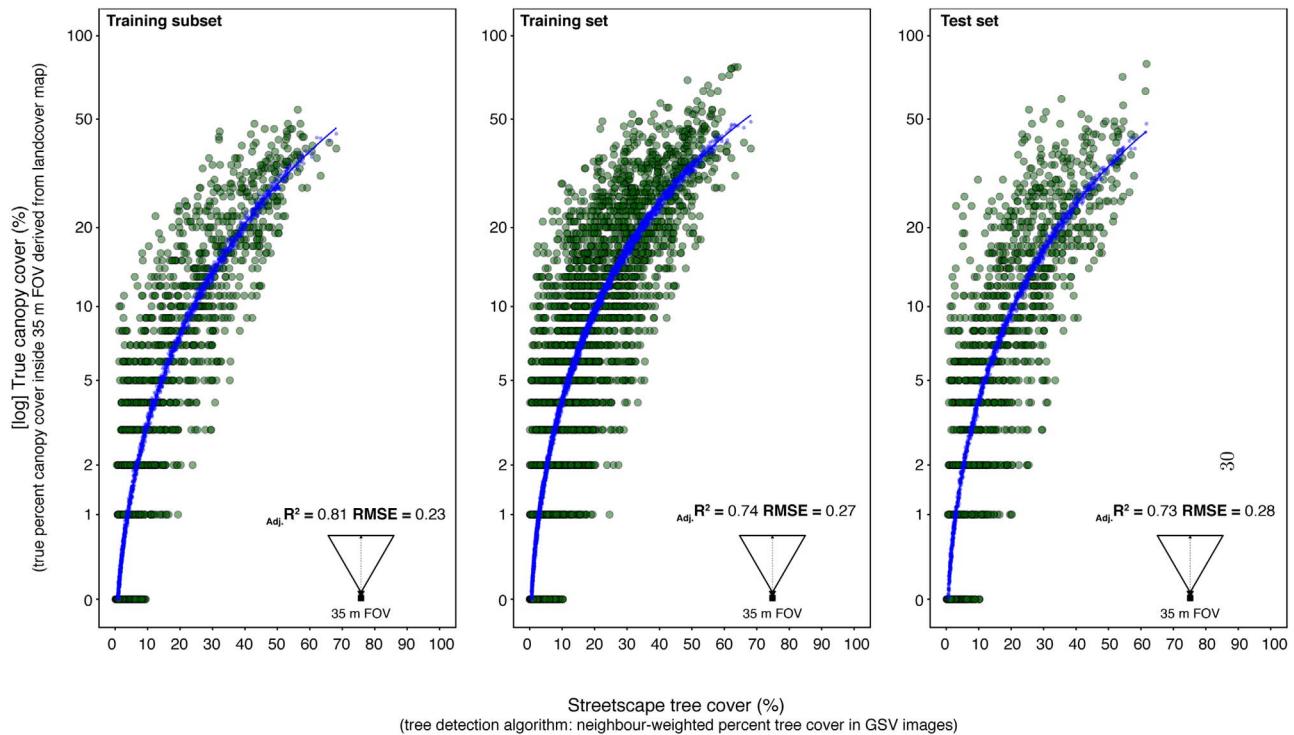


Fig. 3. Relationship between the streetscape tree cover (x-axis) and the true percent tree canopy cover derived from a high resolution landcover map (y-axis displayed on a logarithmic scale) at the 35 m FOV-level for each dataset: the data subset used to learn the weighting factors (left panel), the training set using all data (center panel) and the unseen test data (right panel). Small blue circles are the regression model's predicted values and the blue line is a smoothing line fit to the model's predicted values with a square-root polynomial. The adjusted r-square values and root mean squared-error values for the models are reported in the lower corner of each panel and Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Hand in hand with the growing availability and breadth of open digital data, this study has exemplified how computer vision tools can be applied to quantify patterns of ecological, environmental and urban design importance. Specifically, a novel application of computer vision techniques to digital photographs of a city's streetscapes has shown significant potential to estimate the presence and amount of urban trees at high spatial resolutions with city-wide extent. We may define this quantification of street trees as the perceived tree cover. The multi-step image segmentation method estimated the image area covered by trees in GSV images with high throughput; the algorithm processed about one image per second on an Intel core i7 CPU with 12 cores. The streetscape tree cover method correlated very closely to the classifications performed by a human on a subset of images (see Fig. 2) and this comparison suggested that the algorithm was inclusive of small trees and tree-like plants as well as distant trees in the scene background. It also correlated with a computationally simple, single-feature automated green mask estimator of tree or vegetation presence in an image, though not as well as with the manual pixel masks. This suggests that our computer vision-based method provides a realistic and accurate representation of tree presence and cover while being much less prone to false positives – not everything green is a tree.

A small number of studies have attempted to extrapolate tree presence and coverage from single 2D digital photographs (Li et al., 2015; Peper and McPherson 2003; Schroeder, 1988; Yang et al., 2009) in both urban and natural settings to varying degrees of success. The majority have relied on manual inputs or site-specific conditions to acquire or process images, resulting in low throughput (Peper and McPherson 2003; Schroeder, 1988; Yang et al., 2009). Others have performed automated estimates of vegetation presence, but using non-discriminate, single-feature metrics (e.g., image greenness) and single-image representations of a scene (Li et al., 2015). By applying state-of-

the-art computer vision algorithms, we are able to quantify vegetation represented in images using multiple features and attribute definitive semantic labels to tree-associated pixels.

While the challenge of interpolating a 3D measure of individual tree cover and size from 2D images remains, we attempted to partially correct the proximity-to-camera problem by including information from multiple, neighbouring and overlapping images. The method significantly increased the power of the streetscape tree cover metric to predict true canopy cover. In doing so, we have demonstrated a generalized model of perceived urban tree cover that does not require external inputs beyond the images themselves.

We found that when the streetscape images were aligned parallel to the street direction (*i.e.*, the camera heading was the same as the street heading) the streetscape tree cover predictor operated best relative to the three other road-to-image orientations that we tested. This may be somewhat expected as this orientation generally results in a full view of street trees to either side of the road. In the case of New York City, this result somewhat hampered the analysis because most streets in New York do not have an east-west heading. This is not a limitation of the method in itself however, but rather an artifact of our image sampling scheme. In future applications of this method we can acquire all images at this orientation by adjusting the camera heading when acquiring the images from Google. Alternatively, future applications can make use of GSV panoramas to model the full 360° of the streetscape perceived tree cover.

4.1. Predicting tree canopy cover

The final model estimating true percent canopy cover from our streetscape tree cover method explained a substantial portion of the variation, indicated good prediction accuracy and generalized well to new data. To date, we are not aware of any other urban tree cover mapping methods that correlate so well with state-of-the-art high-

Table 1

Model summary statistics for each final regression model of streetscape tree cover vs. the true percent tree canopy cover. Statistics are shown separately for models using the training, test and a vigorously-iterated training subset (to remove systemic errors) datasets as well as for the preliminary models testing all other FOV-levels. The lower section shows results of a regression analysis using only the un-weighted image tree cover values of the node GSV sampling points at the 35 m FOV-level (i.e., based on a single GSV image and before applying the neighbour-weighted percent tree cover score procedure). The input variable abbreviations are: PTCC, percent tree canopy cover; STC, streetscape tree cover; PG, percent ground; PB, percent building.

| 35 m Field of View | | | |
|----------------------------------|------------------------------------|------------------------------------|------------------------------------|
| Statistic | Training Set | Test Set | Training Subset |
| formula | PTCC in FOV ~ STC + √STC + PG + PB | PTCC in FOV ~ STC + √STC + PG + PB | PTCC in FOV ~ STC + √STC + PG + PB |
| r-square | 0.74 | 0.73 | 0.81 |
| adj.rsquare | 0.74 | 0.73 | 0.81 |
| RMSE | 0.27 | 0.281 | 0.233 |
| MSE | 0.073 | 0.079 | 0.054 |
| df | 3104 | 1328 | 1622 |
| 4-fold cross validation MS | 0.0731 | NA | 0.054 |
| mean fitted canopy cover (log) | 0.822 | 0.81 | 0.641 |
| mean predicted canopy cover(log) | 0.822 | 0.8 | 0.641 |
| SD fitted canopy cover (log) | 0.452 | 0.44 | 0.487 |
| SD predicted canopy cover(log) | 0.526 | 0.52 | 0.539 |

| 15 m Field of View | | | |
|--------------------|------------------------------------|------------------------------------|------------------------------------|
| Statistic | Training Set | 25 m Field of View | 45 m Field of View |
| formula | PTCC in FOV ~ STC + √STC + PG + PB | PTCC in FOV ~ STC + √STC + PG + PB | PTCC in FOV ~ STC + √STC + PG + PB |
| r-square | 0.68 | 0.7 | 0.67 |
| adj.rsquare | 0.68 | 0.7 | 0.67 |
| RMSE | 0.375 | 0.322 | 0.281 |
| MSE | 0.140625 | 0.103684 | 0.078961 |
| df | 3130 | 3147 | 3146 |

| No-neighbour 35 m Field of View ^a | |
|----------------------------------------------|------------------------------------|
| Statistic | Training Subset |
| formula | PTCC in FOV ~ STC + √STC + PG + PB |
| r-square | 0.59 |
| adj.rsquare | 0.59 |
| RMSE | 0.34 |
| MSE | 0.1156 |
| df | 3149 |

^a No-neighbour FOV refers to models using the streetscape tree cover estimation value obtained only from the single GSV image of the node sampling point and, hence, the neighbour-weighting procedure from Eqs. (2) & (3) was not applied.

resolution canopy cover mapping techniques while achieving such high throughput at the city scale. Moreover, this result can be considered conservative since the temporal mismatch between the GSV image data (ca. 2014) and the tree canopy landcover map (ca. 2010) was undoubtedly a source of systemic error cases, and thus unexplained variation. What's more, given the differing perspectives between long-range remote sensing (e.g., satellite imagery and aerial LiDAR) and our street-view photograph-based technique, we would not expect a perfect relationship between the two metrics. Our streetscape metric quantifies the vertical profile of urban trees while landcover mapping uses an overhead view and, thus, sees trees as surfaces or polygons on a horizontal plane.

The results illustrated that this new ground-derived quantification of urban trees can be considered as a *perceived tree cover* estimate rather than canopy cover *per se*. Traditional methods like high-resolution landcover maps are fundamentally fixed as a measure of canopy cover. They are well-suited to compute the coarse-scale distribution and total extent of the tree canopy in a city because they represent trees as a contiguous horizontal layer of leaves covering a given area. Our streetscape metric, on the other hand, is an equally valid measure of tree cover, but is likened to that which is perceived by people at ground-level. As such, it may be better suited to evaluate the spatial variation in tree cover, fine-scale distributions of urban trees and its relationship to other scene features like buildings. Importantly, it presents a quantification of the urban tree cover consistent with the

human perspective; simply put, what people in a city see and experience.

While data limitations in this initial effort prevented an analysis of the full extent of New York city, we demonstrated the applicability of the method to rapidly estimate the total or average amount of perceived tree cover at different city unit sizes. Averaging streetscape tree cover values for different city district-levels correlated well with their true percent tree canopy cover and this correlation increased with district-level unit size (i.e., from the dynamic block level to the borough level). Given its ability for rapid and high throughput, the method represents a promising tool to examine environmental – social and economic patterns within and across cities.

The high resolution and scale that the streetscape tree cover metric achieves will enable a better understanding of the role that trees and vegetation may play in urban dynamics and human health. For example, many studies have evidenced a link between human health benefits and the presence of urban tree cover (Kardan et al., 2015; Nowak et al., 2014; Richardson et al., 2013). However, these studies have been defined by small sample sizes and are limited in their geographic scope to a single city or a few neighbourhoods within them. The ability to quickly quantify urban tree cover for multiple cities concurrently would allow researchers to determine whether the health benefits of urban trees are pervasive or, alternatively, what specific contexts they exist or are maximized in (e.g., biogeographic conditions, local policies and management practices or socioeconomic indicators).

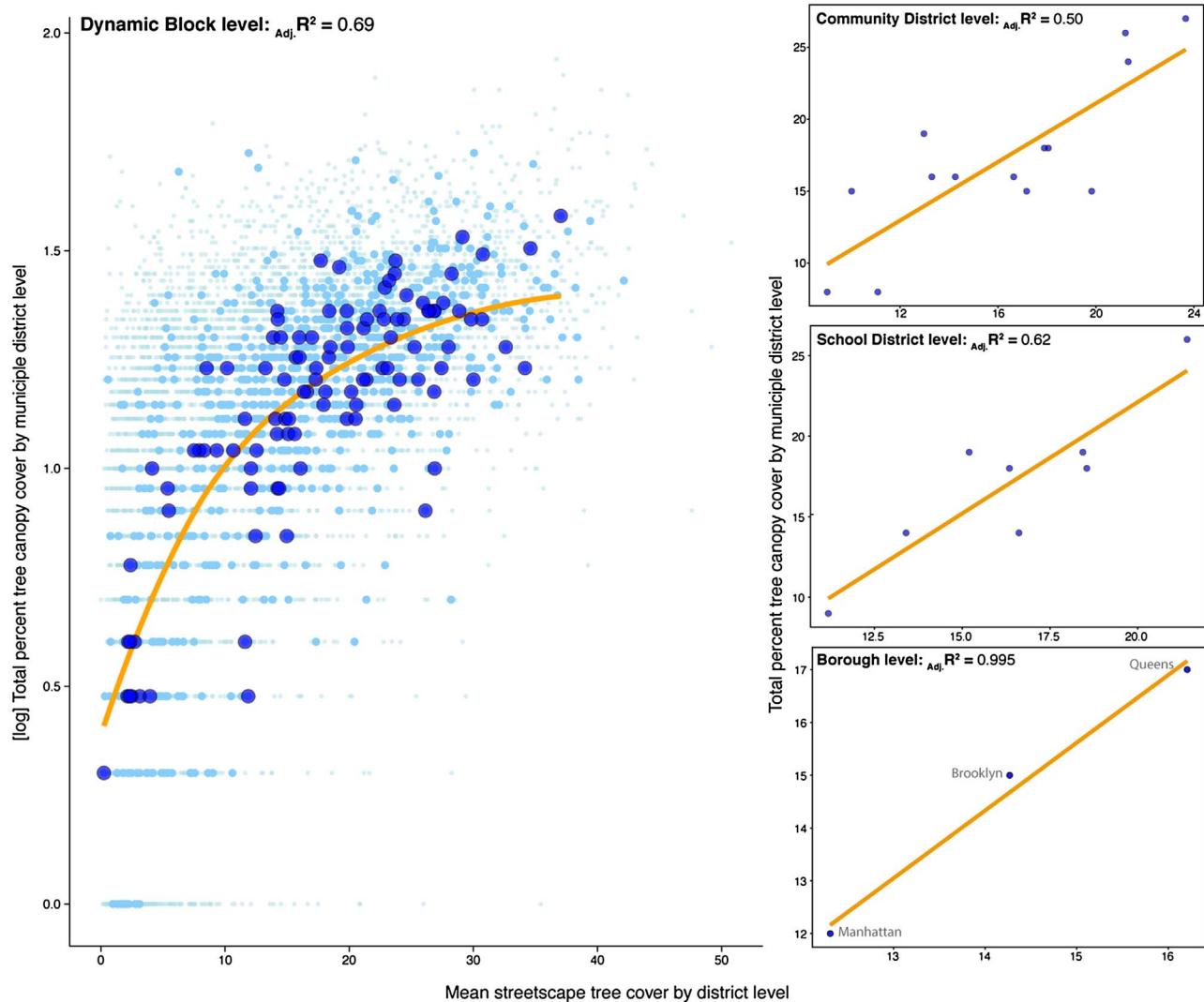


Fig. 4. Relationship between the mean streetscape tree cover (x-axis) and the true percent tree canopy cover (y-axis) at different municipal district-levels of New York City: dynamic census block (large left panel; y-axis is displayed on a logarithmic scale), community district (top right panel), school district (center right panel) and borough (lower right panel). The average unit size of each district-level is reported in Table S2 (SI, Appendix). For the dynamic block district level (left panel), the large blue circles correspond to data points retained after removing those not meeting the minimum count-per-block cutoff value (Table S2; SI, Appendix). The light-blue and smaller circles correspond to data points after increasing the minimum cutoff value (medium-sized, light blue) and all points with no cutoff (smallest, light blue). The regression model fits are shown for each level with the orange lines and the adjusted r-square values reported in the top left panel corners. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

This method permits the analysis of urban tree cover and its relationships with local conditions or social factors at much finer scales than allowed before. Relationships between urban trees and the physical and social components of cities that were previously opaque, such as how income level and social status relates to tree presence and neighbourhood aesthetics, can be investigated in depth.

4.2. Limitations

Systemic errors existed in the data wherein in a range of our streetscape tree cover metric's values were associated with low percent canopy cover values, and in particular 0%. We hypothesize these errors were largely produced by cases in which trees were indeed present, but too small and isolated (e.g., small trees or fine-scale structural features of the tree's shape) to be detected by the landcover mapping methods (MacFaden et al., 2012). In other cases, trees may be present but were occluded in the image FOV by other objects (e.g., buildings or other trees).

The values of the weighting factors changed depending on the training dataset used to compute them, indicating that they are

sensitive to sample size, sample area and the errors associated with both. Though the weighting factor values fluctuated up to 10% with changing sample data, their proportions relative to each other (i.e., neighbours) were consistent. Much through trial and error, we attempted to optimize the weighting factors towards building final predictive models which generalized the best. Future developments of this methodology may benefit from more sophisticated techniques to learn the weighting factors such as non-linear methods.

Regardless of those systemic error cases, the method does not completely compensate for the object proximity-to-camera problem and, hence, explain all of the variation in true tree cover. To fully account for the proximity effect, the model would require additional spatial information (e.g., tree location or, as surrogate inferential features, street width or distance to the sidewalk etc). Over-predicted values of true percent canopy cover were likely due to proximity-to-camera effects that have not been fully accounted for or to the streetscape metric's detection of small vegetation not represented in the landcover mapping methodology. Under-predicted values of true percent canopy cover were likely due to occlusion effects.

We also noticed that tree shadows may, in some cases, be a source of

segmentation error, in that the estimator included some areas of tree shadows as trees in a few cases. In addition, the estimator appeared somewhat sensitive to under- and over-exposed areas of the image, with underexposed areas surrounding a tree being included as tree cover, while overexposed portions of trees being excluded at times.

While a rich and extensive source of data on the world's cities, using GSV images also presents limitations with the most obvious being the coverage is limited to streetscapes. At present, the GSV API has limitations on the number of daily requests and not all sampling years are available since the program began in 2007. What's more, for most cases the streetview images do correspond with days of the year in which the growing season is active in the given geographic area, yet there may exist some locations or cities to which there are exceptions.

5. Conclusions

Quantifying the amount and distribution of trees in cities has been an open challenge due to the fine-scale data required to discern the cluttered and complex spatial heterogeneity defining them. We have addressed this challenge by presenting a novel method of quantifying urban tree cover in city streetscapes using only open source data and software, while achieving very high throughput at the city extent. We validated our streetscape metric by illustrating its ability to estimate percent tree canopy cover with accuracies comparable, and in some cases superior, to established image analysis methods used in landscape ecology.

This new method may be interpreted as a unique measure of urban tree cover – *perceived urban tree cover* – rather than a replacement for high resolution tree canopy cover maps or detailed field surveys. Though not explicitly measured, it contributes inherent information on verticality (tree height) which is not easily obtained through current long-range remote sensing techniques. Taking advantage of the growing library of open source urban image data, the streetscape tree cover metric achieves fine spatial grain at the entire city extent. Importantly, it quantifies urban tree cover from a viewpoint in which urban citizens see and experience the urban landscape, that is the streets.

Acknowledgements

This research was supported by a grant from the Fonds de recherche du Québec – Nature et technologies (FRQNT). The authors would like to thank the valuable insights of Dr. Behrooz Hashemian, Dr. Moe Vazifeh and Dr. Rich Hallett. The authors would also like to thank the New York City Department of Parks and Recreation for sharing datasets on city trees. The authors thank the Amsterdam Institute for Advanced Metropolitan Solutions (AMS), Allianz, Ericsson, Liberty Mutual Research Institute, Philips, the Kuwait-MIT Center for Natural Resources and the Environment, Singapore-MIT Alliance for Research and Technology (SMART), the Société nationale des chemins de fer français (SNCF), UBER, Volkswagen Electronics Research Laboratory, and all the members of the MIT Senseable City Lab Consortium for supporting this research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.landurbplan.2017.05.010>.

References

Dickinson, J. L., Zuckerberg, B., & David, B. N. (2010). Citizen Science as an Ecological

- Research Tool: Challenges and Benefits. *Annual Review of Ecology and Systematics*, 41, 149–172. <http://www.jstor.org/stable/27896218>.
- Felzenswalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-Based image segmentation. *International Journal of Computer Vision*, 59(2), 167–181. <http://dx.doi.org/10.1023/B:VISI.0000022288.19776.77>.
- Fuller, R. A., & Gaston, K. J. (2009). The scaling of green space coverage in European cities. *Biology Letters*, 5(3), 352–355. <http://dx.doi.org/10.1098/rsbl.2009.0010>.
- Glaeser, E. L., Kominers, S. D., Luca, M., & Naik, N. (2015). *Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life*. <http://dx.doi.org/10.3386/w21778> NBER Working Paper.
- Google Inc (2014). *Google street view image API*. <https://developers.google.com/maps/documentation/streetview/intro?hl=en>.
- Hoiem, D., Efros, A. A., & Hebert, M. (2005). Geometric context from a single image. *Tenth IEEE international conference on computer vision (ICCV'05)*. vol. 1, (pp. 654–661). <http://dx.doi.org/10.1109/ICCV.2005.107>.
- Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., et al. (2007). Completion of the 2001 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, 73(4).
- Kardan, O., Gozdyra, P., Misic, B., Moola, F., Palmer, L. J., Paus, T., et al. (2015). Neighborhood greenspace and health in a large urban center. *Scientific Reports*, 5(January), <http://dx.doi.org/10.1038/srep11610>.
- Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., John Kress, W., Lopez, I. C., et al. (2012). A computer vision system for automatic plant species identification what plant species is this? *Eccv*, 1–14. http://dx.doi.org/10.1007/978-3-642-33709-3_36.
- Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, 14(3), 675–685. <http://dx.doi.org/10.1016/j.ufug.2015.06.006>.
- Lothian, A. (1999). Landscape and the philosophy of aesthetics: Is landscape quality inherent in the landscape or in the eye of the beholder? *Landscape and Urban Planning*, 44(4), 177–198. [http://dx.doi.org/10.1016/S0169-2046\(99\)00019-5](http://dx.doi.org/10.1016/S0169-2046(99)00019-5).
- Lovasi, G. S., Quinn, J. W., Neckerman, K. M., Perzanowski, M. S., & Rundle, A. (2008). Children living in areas with more street trees have lower prevalence of asthma. *Journal of Epidemiology and Community Health*, 62(7), 647–649. <http://dx.doi.org/10.1136/jech.2007.071894>.
- MATLAB (2015). *Version 8.5.0.197613 (R2015a)*. Natick, Massachusetts: The MathWorks Inc.
- MacFaden, S. W., O'Neil-Dunne, J. P. M., Royar, A. R., Lu, J. W. T., & Rundle, A. G. (2012). High-resolution tree canopy mapping for New York City using LiDAR and object-based image analysis. *Journal of Applied Remote Sensing*, 6(1), <http://dx.doi.org/10.1117/1.JRS.6.063567>.
- McPherson, G., Nowak, D., Heisler, G., Grimmond, S., Souch, C., Grant, R., et al. (1997). Quantifying urban forest structure, function, and value: The Chicago Urban Forest Climate Project. *Urban Ecosystems*, 49–61. <http://dx.doi.org/10.1023/A:1014350822458>.
- Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers Electronics in Agriculture*, 63(2), 282–293.
- Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore – Predicting the perceived safety of one million streetscapes. *2014 IEEE conference on computer vision and pattern recognition workshops*, 793–799. <http://dx.doi.org/10.1109/CVPRW.2014.121>.
- Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L., & Hidalgo, C. A. (2015). Do people shape cities, or do cities shape people? The Co-evolution of physical, social, and economic change in five major U.S. cities. *National bureau of economic research working paper series No. 21620*. <http://dx.doi.org/10.3386/w21620>.
- Nowak, D. J., Hirabayashi, S., Bodine, A., & Greenfield, E. (2014). Tree and forest effects on air quality and human health in the United States. *Environmental Pollution*, 193, 119–129.
- Peper, P. J., & McPherson, G. E. (2003). Evaluation of four methods for estimating leaf area of isolated trees. *Urban Forestry & Urban Greening*, 2(1), 19–29. <http://dx.doi.org/10.1078/1618-8667-00020>.
- Proulx, R., Roca, I. T., Cuadra, F. S., Seiferling, I., & Wirth, C. (2014). A novel photographic approach for monitoring the structural heterogeneity and diversity of grassland ecosystems. *Journal of Plant Ecology*, 7(6), 518–525. <http://dx.doi.org/10.1093/jpe/rtt065>.
- R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Richardson, E. A., Pearce, J., Mitchell, R., & Kingham, S. (2013). Role of physical activity in the relationship between urban green space and health. *Public Health*, 127(4), 318–324. <http://dx.doi.org/10.1016/j.puhe.2013.01.004>.
- Schroeder, H. W. (1986). Estimating park tree densities to maximize landscape esthetics. *Journal of Environmental Management*, 23(4), 325–333. <http://cat.inist.fr/?aModele=afficheN& & }cpsidt=8381648>.
- Schroeder, H. W. (1988). Visual impact of hillside development: Comparison of measurements derived from aerial and ground-level photographs. *Landscape and Urban Planning*, 15(1-2), 119–126. [http://dx.doi.org/10.1016/0169-2046\(88\)90020-5](http://dx.doi.org/10.1016/0169-2046(88)90020-5).
- Thayer, R. L., & Atwood, B. G. (1978). Plants, complexity, and pleasure in urban and suburban environments. *Environmental Psychology and Nonverbal Behavior*, 3(2), 67–76. <http://dx.doi.org/10.1007/BF01135604>.
- Yang, J., Zhao, L., McBride, J., & Gong, P. (2009). Can you see green? Assessing the visibility of urban forests in cities. *Landscape and Urban Planning*, 91(2), 97–104. <http://dx.doi.org/10.1016/j.landurbplan.2008.12.004>.