

# Quantifying urban forest structure with open-access remote sensing data sets

Oliver Baines <sup>1</sup>, Phil Wilkes <sup>1,2\*</sup>, and Mathias Disney <sup>1,2</sup>,

<sup>1</sup> *Department of Geography, University College London, Gower Street, London, WC1E 6BT*

<sup>2</sup> *NERC National Centre for Earth Observation*

*\* Correspondence: p.wilkes@ucl.ac.uk*

---

## Abstract

Future cities are set to face ever increasing population and climate pressures, ecosystem services offered by urban forests have been recognised as providing significant mitigation for these pressures. Therefore, the ability to accurately quantify the extent and structure of urban forests, across large and highly dynamic cities, is vital for determining the value of services provided and to assess the effectiveness of policy to promote these important assets. Current inventory methods used in urban forestry are mostly reliant on plot networks measuring a range of structural and demographic metrics; however, limited sampling (spatially and temporally) cannot fully capture the dynamics and spatial heterogeneity of the urban matrix. The rapid increase in the availability of open-access remote sensing data and processing tools offers an opportunity for monitoring and assessment of urban forest structure that is synoptic and at high spatial and temporal resolutions. Here we present a framework to estimate urban forest structure that uses open-access data and software, is robust to differences in data sources, is reproducible and is transferable between cities. The workflow is demonstrated by estimating three metrics of 3D forest structure (canopy cover, canopy height and tree density) across the Greater London area (1,577 km<sup>2</sup>). Random Forest was trained with open-access airborne LiDAR or iTree Eco inventory data, with predictor variables derived from Sentinel 2, climatic and topography data sets. Output were maps of forest structure at 100 m and 20 m resolution. Results indicate that forest structure can be accurately estimated across large urban areas; Greater London has a mean canopy cover of ~16.5% (RMSE 11 - 17%), mean canopy height of 8.1 - 15.0 m (RMSE 4.9 - 6.2 m) and is home to ~4.6M large trees (projected crown area >10 m<sup>2</sup>). Transferability to other cities is demonstrated using the UK city of Southampton,

where estimates were generated from local and Greater London training data sets indicating application beyond geographic domains is feasible. The methods presented here can augment existing inventory practices and give city planners, urban forest managers and greenspace advocates across the globe tools to generate consistent and timely information to help assess and value urban forests.

*Keywords:* urban forest structure, open-access, remote sensing, airborne LiDAR, iTree Eco, Sentinel 2

---

## 1. Introduction

Urban forests are now recognised as a key asset for maintaining livable cities in the face of rapidly expanding populations and climate change [1, 2, 3]. To this end, governments are setting ambitious targets to increase urban forest cover at all scales [4]; for example, Melbourne is aiming for a 20% increase in canopy cover by 2040 [5] and Vancouver aims to increase canopy cover by ~0.5% per annum until 2030 [6]. To assess the effectiveness of policy, a comprehensive understanding of contemporary urban forest structure is required, as well as the ability to monitor change through repeatable measurements and methods.

Assessment has so far been achieved using inventory techniques borrowed from traditional forestry, where a suite of metrics are recorded across a plot network and aspatial statistics reported. Recently, inventory protocols have been developed specifically for urban areas with the aim of standardising monitoring and assessment of forest structure and demographics [7]. Developed by the US Department of Agriculture (USDA) Forestry Service, the iTree Eco protocol [8] is one example that has seen widespread adoption in cities globally; for example, London [9], New York [10] and Barcelona [11]. An output of iTree Eco analysis is a monetary value for ecosystem services, providing valuable information for land managers as well as summarising intangible benefits to a wider audience [9]. Although inventory measurements are locally comprehensive, they can be sparse in a heterogeneous urban landscape, laborious and time-consuming to measure [7, 12] and reliant on a volunteer workforce [4] as well as often being restricted by land ownership [13].

Remote sensing techniques offer a new approach for assessing urban forests that have already been widely adopted for national inventories [14, 15]. These techniques encompass a wide range of measurement technologies (active and passive) and scales (sub-metre to kilometre); providing spatially

30 and temporally synoptic data across government and cadastral boundaries.  
31 Remote sensing of urban forest structure poses particular challenges however  
32 [16], including: a spatially heterogeneous and fragmented patchwork of land  
33 cover and forest types, from isolated trees to dense forest; high tree species  
34 diversity including non-natives; heavily and regularly managed trees (e.g.  
35 pollarding); and occlusion by tall buildings [12, 17]. Despite this, recent ad-  
36 vances in remote sensing technologies and processing methods have enabled  
37 accurate assessments of urban forest structure and composition; Tigges and  
38 Lakes [18] and Li et al. [19] provide comprehensive reviews of remote sensing  
39 of urban forest structure.

40 Canopy cover metrics (and their synonyms) are a 2D estimate of forest  
41 extent which is often reported in urban forest studies. A common remote  
42 sensing approach to assessing canopy cover is random point sampling of  
43 aerial or satellite imagery [20, 21]. For example, the USDA Forest Service  
44 iTree Canopy tool [22] uses Google Earth imagery as a base map to enable  
45 a user to determine whether a point is canopy or not (typically 500 - 1000  
46 sample points). More complex statistical techniques, such as regression,  
47 supervised classification and machine learning, have also been applied to  
48 estimate canopy cover [23, 24]. Owing to a greater number of sample points,  
49 application of remote sensing techniques has reduced uncertainty in urban  
50 canopy cover estimates [21].

51 Although canopy cover is a relatively simple and useful metric to derive,  
52 it is just one of a suite of potential metrics for assessing the extent, health  
53 and demographics of urban forests that can be derived from remote sensing.  
54 Other structural metrics include canopy height, leaf area index (LAI) [25]  
55 and projected crown area [17] as well as those derived from canopy reflect-  
56 ance e.g. tree species [26, 27]. A combination of structural and spectral  
57 metrics have also been used, for example, to segment individual trees [28].  
58 Estimating urban forest extent using the UN Forest and Agriculture defin-  
59 ition for “*other forested areas*” [29] would require an assessment of canopy  
60 height as well as canopy cover.

61 3D information of canopy structure produced by LiDAR (light detec-  
62 tion and ranging) measurements can surpass the information from passive  
63 instruments for assessing forest structure. Airborne LiDAR has been suc-  
64 cessfully used to estimate biophysical parameters such as canopy height and  
65 cover from either the original point cloud data or rasterised products e.g. a  
66 canopy height map. When applied in urban areas, the use of point cloud  
67 data allows for the removal of tall buildings and other infrastructure before  
68 canopy metrics are derived [30, 17]. Until the late 2000s, large area acqui-  
69 sitions of ALS over urban areas were uncommon or not publicly accessible;

70 in this case, sub-area acquisition have been used to train predictive models  
71 of canopy cover [24] and above ground biomass (AGB) [31]. Where wall-to-  
72 wall data are available, airborne LiDAR has been used to estimate leaf area  
73 index [25] and AGB [17].

74 Since the late 2000s, many remote sensing data have moved from propri-  
75 etary to open-access distribution models, driving an uptick in new applic-  
76 ations [32]. Availability of open-access data sets is not limited to satellite  
77 remote sensing products, a number of countries have opened their archives  
78 of aerial imagery and LiDAR data. For example, the UK Environment  
79 Agency have made available LiDAR covering 72% of England [33]; similar  
80 data sets are available for Finland, Belgium, Spain, the United States and  
81 New Zealand to name but a few. However, not all LiDAR data are equal:  
82 for example, data may be pre-filtered to remove vegetation if the purpose of  
83 the data processing is flood mapping [17] or acquisition specifications (alti-  
84 tude, pulse density, registration accuracy etc.) may differ between surveys.  
85 New analysis methods and workflows that use open-access data need to be  
86 robust to these differences.

87 A potential barrier to adoption of remote sensing techniques has been the  
88 (perceived) level of specialist knowledge required. Recently however, new  
89 methods and workflows e.g. through web-based applications, are allowing  
90 people without specialist knowledge to access and manipulate these data  
91 much more easily. For example, Google Earth Engine (GEE) is a cloud-  
92 based platform, accessed through a web-based programming interface, which  
93 allows technical and non-technical users to process remote sensing data sets  
94 [34]. Applications of GEE have so far included mapping of urban [35] and  
95 woody vegetation [36] extent to name but a few.

96 The new emphasis on urban forest and their importance for maintaining  
97 liveable cities coupled with a new era of open-access data and processing  
98 workflows makes remote sensing the ideal monitoring tool for these import-  
99 ant assets. Here we present a framework that utilises open-access LiDAR and  
100 satellite data to quantify urban forest structure, specifically canopy height,  
101 cover and tree density. We demonstrate this framework using Greater Lon-  
102 don (Figure 1), a city with a long history of urban forestry and that has  
103 been recently designated the first *National Park City*. We also demonstrate  
104 the approach is transferable to other areas by performing an assessment of  
105 Southampton, a smaller city in the south of the UK.



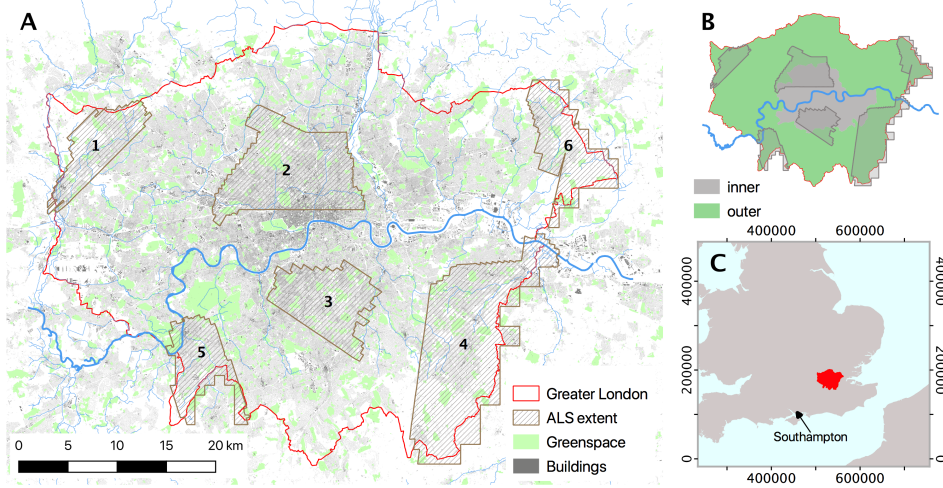


Figure 1: Map of study area. (A) Greater London with LiDAR extent overlaid (zones are numbered) (B) Inner and Outer London and (C) the location of Greater London in the wider context. Contains data from Rae [37] and OS data © Crown copyright and database right (2019)

## 2. Materials and Methods

The aim of this research is to present a framework for predicting wall-to-wall urban forest structure that is transferable to any urban area. This is achieved by first training a predictive model (in this case Random Forest [38]) with forest structure metrics and a set of synoptic predictor variables, before applying this wall-to-wall across the region of interest (Figure 2). To make the framework more widely applicable, data sets (Table 1) and tools (see Appendix A) are open-access. The framework is designed to be modular; for example, forest structure metrics have been derived from both airborne LiDAR (ALS) and iTree Eco inventory data.

Three metrics of structure have been modelled; Canopy Cover  $CC$ , Canopy Height  $CH$  and tree density  $N$ .  $CC$  is widely-used as a proxy for assessing urban forest extent,  $CH$  is required to more accurately estimate forest extent and can also be used to derive secondary metrics (e.g. AGB),  $N$  can be used to determine the distribution and demographics of tree populations. Definitions of these metrics can vary between studies, making inter-comparison difficult; for example  $CC$  may or may not include the shrub layer [13]. Below, metrics are explicitly defined dependent upon measure-

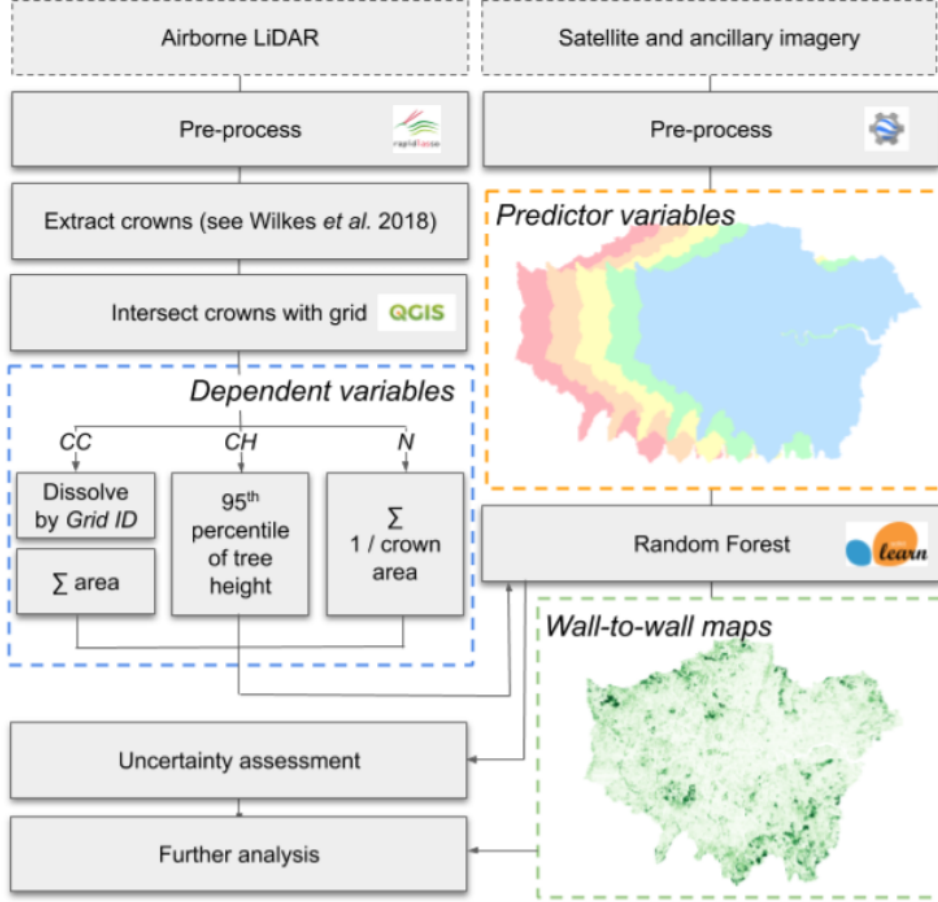


Figure 2: Schematic of framework showing the processing of airborne LiDAR to create a training data set.

ment technique, with the aim of making these transferable and comparable across sites and studies. Tree attributes were aggregated at two grid sizes; 100 m and 20 m grids where the former is equivalent to 1 ha blocks and the latter is the minimum resolution of the predictor variables.

### 2.1. Study Area

Greater London (Figure 1) covers an area of 1,572km<sup>2</sup> and is home to 8.8M people [39] and 8.4M trees (diameter at breast height  $dbh \geq 0.07$  m)[9]. Land use in the region is a matrix of recreational green space (38.2%),

residential (32.6%), transport infrastructure (14.1%) and industrial (4.7%) [40]. Greater London can be split into two broad sub-regions: Inner (1.6 M trees) and Outer London (6.8 M trees, Figure 1B) [9]. Land use in Inner London is primarily urban and industrial (but includes several large parks) and Outer London is urban to open countryside including large extents of forest and farm land.

Greater London has a range of forest types, from isolated single trees to 2,500 ha of ancient (i.e. persisted since at least 1600) woodland [41], with an estimated amenity value of £43.3 billion [9]. Canopy cover estimates range from 14% [9] to 21.9% [42] dependent on method and definition. There are a number of sources of tree and tree metric information for Greater London, including the *Local Authority Maintained Trees* database of 700,000 trees [43]. However, none are complete even locally (e.g. often restricted to highways only) highlighting the requirement for a synoptic method which is not restricted by access to private land.

## 2.2. Remote sensing and inventory data

### 2.2.1. Airborne LiDAR data

Airborne LiDAR data (aka Airborne Laser Scanning or ALS) was captured over an area of ~500 km<sup>2</sup>, split into 6 zones, by the UK Environment Agency in February - March of 2015 and 2016 (Figure 1A). 1 km tiles were downloaded in .laz format from the UK Government's data portal (Open Government Licence v3.0). Point cloud data was pre-processed using LAStools [44] to ground normalise, remove flight line overage and extract points that were attributed as having multiple returns. Tree crown extraction then followed the method of Wilkes et al. [17]. Identified crown envelopes where crown height was <2 m or >40 m or where projected crown were <10 m<sup>2</sup> were discarded. A total of 1,590,128 individual trees were identified.

Two regular grids (100 m and 20 m) were intersected with the vector crown layer where the origins of the grids matched that of the predictor data (see below). Some crowns intersected more than one grid cell and were therefore split into sub-crown areas. Canopy cover  $CC_{ALS}$  was computed as the sum of projected crown area (ignoring within crown gaps) where overlapping crown polygons were dissolved, divided by grid square area. Canopy height  $CH_{ALS}$  was calculated from the 95th percentile of crown heights per grid square. Tree density  $N_{ALS}$  was estimated as the sum of the proportions of crowns in a grid square. Total number of trees  $\sum N$  was estimated by summing  $N_{ALS}$ .

Table 1: Predictor variables derived from open-access data sets available in GEE.

Source	20 m	100 m
<b>Sentinel Level 2A (bands 1 - 9, 11, 12)</b>		
Surface reflectance	✓	✓
NDVI	✓	✓
Mean (3 x 3 window)	✓	✓
Mean (5 x 5 window)	✓	
Texture (3 x 3 window incl. NDVI)	✓	✓
Texture (5 x 5 window incl. NDVI)	✓	
Intra-pixel texture (incl. NDVI)		✓
NDVI temporal variance	✓	✓
<b>WorldClim BIO Variables V1</b>	✓	✓
<b>SRTM</b>		
Elevation	✓	✓
Slope	✓	✓
Aspect	✓	✓

### 2.2.2. Predictor variables

A suite of predictor variables were selected for this analysis based upon previous research [45, 46]; a list of predictor variables is provided in Table 1. These were accessed and pre-processed using GEE to a common format, projection (OSGB 1936) and spatial resolution [34]. A Sentinel 2 Level-2A cloud-free mosaic was computed from images captured between 1/5/2018 - 31/8/2018 (i.e. leaf-on). From this a water mask was computed using the *MNDWI* method where a threshold of 0.1 was applied to mask water [47]. Normalised Difference Vegetation Index (NDVI) variables and texture metrics were derived from the Sentinel mosaic. NDVI temporal variance was computed from the cloud free image stack used to create the mosaic. Intra-pixel variance was computed for 100 m pixels using the 20 m input data sets, where standard deviation of pixel values in the larger grid cell were computed.

Although Random Forest is robust to colinear predictor variables, these can lead to anomalies in variable importance metrics. For a more parsimonious model, co-linear predictor variables ( $R^2 > 0.9$ ) were identified and, by computing pair-wise correlation coefficients with  $CC_{ALS}$ , the least correlated variable was discarded. This reduced the initial set of predictor variables from 74 and 85 to 47 and 49 for 100 m and 20 m analysis respectively.

### 2.2.3. *iTree inventory data*

Inventory data were captured as part of the 2014 London iTree Eco project [9] where a total of 721 0.04 ha circular plots were measured. Plot-wise analogues of the ALS structure metrics were derived from the iTree Eco database where “Percentage Tree Cover”  $CC_{iTree}$ , 95th percentile of measured tree height  $CH_{iTree}$  and a count of all trees within a plot  $N_{iTree}$  were used for  $CC$ ,  $CH$  and  $N$  respectively. Trees with a projected crown area  $<10 \text{ m}^2$  (calculated from the major and minor axis measurements and corrected for % crown missing) were excluded inline with the ALS crown segmentation method [17]. This resulted in an estimated 5,167,000 “large trees” (corrected for % of plot area measured) in the Greater London area ( $32.8 \text{ trees ha}^{-1}$ ); removal of small trees from the database increased median  $dbh$  from 0.19 m to 0.25 m.

To directly compare  $CH_{ALS}$  and  $N_{ALS}$  with iTree Eco data, plot coordinates were buffered by 11.3 m (equivalent to iTree plot radii) and intersected with the ALS derived crown envelope centroids (assumed as germination point). To compare  $CC_{iTree}$  and  $CC_{ALS}$  values, ALS derived crown envelopes were intersected with iTree Eco plot boundaries and the clipped crown area was summed i.e. trees whose germination point was potentially outside the plot were included. To compare Random Forest output with iTree Eco data as well as train a Random Forest with iTree Eco data, grid squares that intersected iTree Eco plots were extracted and mean values for the predictor variables or Random Forest output were computed.

### 2.3. *Random Forest*

Random Forest [38] was chosen for this analysis for a number of reasons; (1) robustness to noise within training data [48] (2) scalability (e.g. [49]) (3) easy to parameterise and (4) widely available e.g. implementations in R, Python and GEE. With regard to remote sensing of forest structure, Random Forest has been used to predict forest cover [46], canopy height [49, 45, 50] and AGB [51]

Here, Random Forest was run as a regressor, using the Python scikit-learn implementation [52], predicting continuous variables of canopy cover  $CC_{RF}$ , canopy height  $CH_{RF}$  and tree density  $N_{RF}$ . Preliminary analysis (unpublished MSc thesis) suggested error stabilised at 50 trees, tree maximum depth was set to 20 to avoid overfitting in lower branches and  $N/3$  variables were selected at each node where  $N$  is the size of the variable pool.

It is implicit that spatial autocorrelation will underestimate error in predictive modelling; to avoid this, a semivaraince analysis using Sentinel 2

Band 3 surface reflectance data was completed. This indicated that at resolutions of 100 m and 20 m variance reached a sill at approximately 600 m and 200 m respectively. Regular grids were therefore imposed on the full resolution gridded data set to subset a spatially uncorrelated cohort; this resulted in a training pool of 7,086 and 110,780 points for 100 m and 20 m resolutions respectively. Training data sets of 5,000 points were sampled (with replacement) from these subsets. Training samples were weighted to be representative of Inner and Outer London total area (22:78) as Inner London was over represented in the ALS data. To compute model accuracy, a test cohort of 2,000 and 5,000 points for 100 m and 20 m respectively was then sampled from remaining uncorrelated points. When training a Random Forest with iTree Eco data, a test cohort of 500 points was sampled from the 741 plots; the remainder were used for testing. Mean and standard deviation values were derived from 25 iterations of Random Forest where for each iteration a new training subset was selected.

Random Forest has been shown to display a systematic error, known as “regression to the mean”, when predicting continuous variables [53, 45]. This results in over and under estimation of smaller and larger values respectively, where model residuals are highly correlated with the independent variable  $X$ . Correcting this error is particularly important when estimating cumulative variables, such as  $\sum N$ , where larger values are disproportionately more important, or to accurately represent the tails of  $X$ . To correct for systematic error, linear regression of model residual ( $X_{RF} - X$ ) and  $X$  as dependent and independent variables respectively was used to compute a correction function. This was then applied to the modelled output and is denoted as  $RF + C$ . Applying the correction can in some cases produce unrealistic values e.g.  $CC_{RF} < 0$ ; unrealistic values were replaced with minimum or maximum values accordingly.

#### 2.4. Model transferability

Applying a model outside the domain of the training data is likely to decrease model suitability. However, as urban areas tend to be ‘artificial’ in terms of forest structure (e.g. managed trees, regular spaced planting, similar species composition), commonality between unconnected urban areas may exist (although see [54]). As suitable training data may not always be available, we test the potential transferability of a Random Forest model between geographic areas. Firstly, this is tested within Greater London where a single “zone” (Figure 1) is used to train the model, this is then applied to the remaining five areas. A second experiment tests transferability to areas further afield, to do this the UK city of Southampton was chosen

Table 2: Results for estimating forest structure at 100 m resolution where **ALS** values are derived from training data, **RF<sub>ALS</sub>** are results from ALS trained Random Forest, **RF+C<sub>ALS</sub>** as before including linear correction and **RF+C<sub>GL</sub>** as before but applied across the Greater London area. *std* is mean within-sample variation unless <sup>B</sup> then standard deviation of the mean where  $N=25$ .

	<b>ALS</b>		<b>RF<sub>ALS</sub></b>			<b>RF+C<sub>ALS</sub></b>			<b>RF+C<sub>GL</sub></b>	
	mean	std	mean	std	rmse	mean	std	rmse	mean	std <sup>B</sup>
<i>CC</i> (%)	17.3	16.77	17.8	12.56	10.90	17.0	17.45	11.52	16.2	0.22
<i>CH</i> (m)	16.6	4.67	16.5	2.33	3.91	16.6	3.96	4.09	15.0	0.10
<i>N</i> (trees ha <sup>-1</sup> )	28.6	21.96	30.2	13.63	18.90	28.0	21.75	20.50	28.4	0.42
$\Sigma N$	4,493,236	100,140 <sup>B</sup>	4,740,741	58,029 <sup>B</sup>		4,397,041	89,793 <sup>B</sup>		4,531,016	67,225

(Figure 1C). Southampton is situated on the south coast ~130 km south west of London and has a population of 250,000 residents. A 2017 iTree Eco report for the city ( $N = 414$  plots) suggests that Southampton has a *CC* of 18.5% and a total of 267,500 trees [55]. ALS was captured over Southampton in March 2016 and training data was derived from trees segmented using the method described in Section 2.2.1. Random Forest modelled canopy structure metrics were generated and compared from the locally derived training data as well as from models trained on the Greater London. Again, 25 iterations of Random Forest were run to generate mean and standard deviation values.

### 3. Results

For the Greater London area,  $CC_{RF}$  is  $16.2\% \pm 0.22\%$  and  $16.5\% \pm 0.37\%$  for analysis at 100 m and 20 m respectively, mean  $CH_{RF}$  is  $15.0 \text{ m} \pm 0.1 \text{ m}$  and  $8.1 \text{ m} \pm 0.2 \text{ m}$  for 100 m and 20 m resolutions respectively and  $N_{RF}$  is ~29 trees ha<sup>-1</sup> for both resolutions with a total of between 4.5M - 4.6M large trees (Tables 2 and 3).  $CC_{RF}$  is slightly higher than the 14% “Tree Cover” value reported in the London iTree Eco survey [9] and the value derived from subsampling the iTree Eco data (Table 4). The London Tree Officers Association [42] estimated cover of 21.9 % for 2010 and the Greater London Authority [43] estimated cover of 19.5% for 2011; however, both estimates were derived from a supervised classification that included small trees and shrubs. For UK urban areas nationally, a mean *CC* of 15.8% is reported [4].  $CH_{RF}$  estimates vary between analysis resolutions, this indicates a non-random tree distribution where a greater number of treeless pixels are identified at 20 m that decreases mean  $CH_{RF}$ . Tree density estimates are broadly similar between techniques (28-29 trees ha<sup>-1</sup>), Random Forest total number of trees is ~10% less than estimated from the iTree Eco database.

Table 3: Results for estimating forest structure at 20 m resolution. See Table 2 for notes. Values in brackets are normalised to 1 ha for comparison with Table 2

	<b>ALS</b>		<b>RF<sub>ALS</sub></b>		rmse	<b>RF+C<sub>ALS</sub></b>		rmse	<b>RF+C<sub>GLA</sub></b>	
	mean	std	mean	std		mean	std		mean	std <sup>B</sup>
<i>CC</i> (%)	15.8	24.42	17.3	18.17	16.31	16.3	25.72	17.81	16.5	0.37
<i>CH</i> (m)	8.4	8.16	8.8	5.12	5.74	8.6	8.00	6.20	8.1	0.15
<i>N</i> (trees px <sup>-1</sup> )	1.0 (25.6)	1.56	1.2 (29.2)	0.89	1.44	1.1 (27.0)	1.51	1.56	1.2 (29.2)	0.03
$\sum N$	4,027,284	92,466 <sup>B</sup>	4,588,941	101,080 <sup>B</sup>		4,243,535	132,325 <sup>B</sup>		4,601,864	131,381

Table 4: Results for estimating forest structure at 20 m resolution where **iTree** values are derived from training data, **RF<sub>iTree</sub>** are results from iTree Eco trained Random Forest and **RF<sub>GLA</sub>** are as before but applied across the Greater London area. *std* is mean within-sample variation unless <sup>B</sup> then standard deviation of the mean where *N*=25. Values in brackets are normalised to 1 ha for comparison with Table 2.

	<b>iTree</b>		<b>RF<sub>iTree</sub></b>		rmse	<b>RF<sub>GLA</sub></b>	
	mean	std	mean	std		mean	std <sup>B</sup>
<i>CC</i> (%)	13.1	22.06	14.4	12.47	18.79	15.9	0.90
<i>CH</i> (m)	4.9	6.78	5.3	3.12	6.31	5.8	0.29
<i>N</i> (trees px <sup>-1</sup> )	1.3 (33.4)	2.47	1.5 (36.6)	1.18	2.24	1.6	0.10
$\sum N$	5,251,702	942,628 <sup>B</sup>	5,758,010	664,951 <sup>B</sup>		6,331,778	389,787

Random Forest tends to overestimate values of *CC*, *CH* and *N* compared to ALS estimated values (Figure 3). A suggested reason for this is that, with the exception of *CH*<sub>100</sub>, training data values are left-skewed (Figure 3). However, Random Forest’s “regression to the mean” tendency reduces skewness, leading to over and underestimated lower and higher values respectively [53, 45]. This tendency also decreases the standard deviation of the output compared to training values i.e. decreased kurtosis (Table 3 and Figure 3). Applying a correction function to Random Forest computed values (*RF* + *C*) increases kurtosis and standard deviation to closely recreate distributions of training data (Figure 3) as well reducing the correlation between the dependent variable and model residuals (Appendix B). However, correction functions are indiscriminate and correctly computed values can be wrongly scaled, which results in an increase in overall error (Tables 2 and 3 and Appendix B).

A comparison of iTree and ALS derived variables (*N* plots = 196) suggests a generally good agreement (Figure 4 top); RMSE is 13.8%, 4.9 m, and 2.9 trees for *CC*, *CH* and *N* respectively. Of the 196 plots 60% and 53% of plots were treeless for iTree Eco and ALS plots respectively. *CC*<sub>ALS</sub> > *CC*<sub>iTree</sub> and *CH*<sub>ALS</sub> > *CH*<sub>iTree</sub>, although particularly for *CC*, measurement techniques were different i.e. iTree Eco uses a visual assessment to determine *CC*. Discrepancies in *CH* were caused by the ambiguity



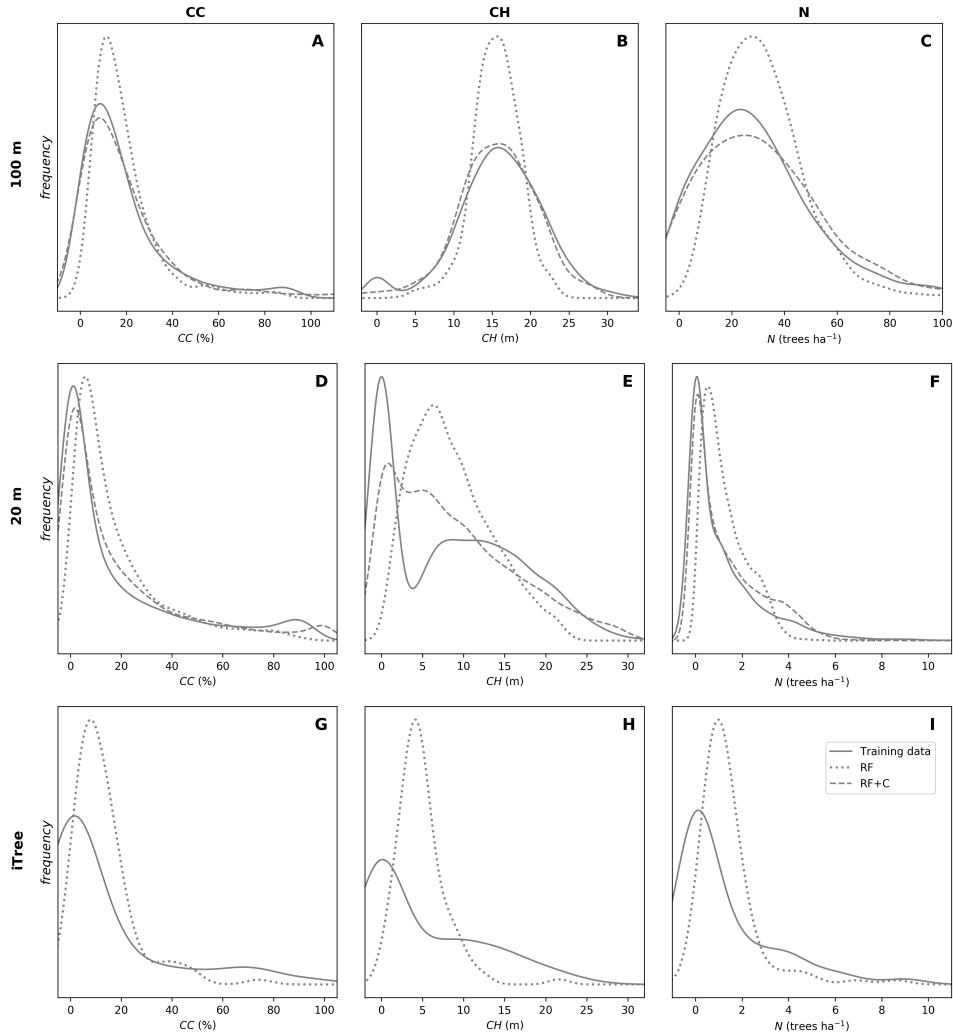


Figure 3: Frequency distributions of  $CC$  (left),  $CH$  (middle) and  $N$  (right) for variables derived at 100 m resolution (A-C), 20 m resolution (D-F) and using iTree Eco data (G-I). Each subplot plots training data and distributions produced from Random Forest (RF) and Random Forest including correction of systematic error. Kernel Density Estimation was applied to smooth trends, hence values  $< 0$ .

314 of germination point i.e. ALS derived crown envelope polygon centroids  
 315 incorrectly inside or outside plot boundaries. ALS tends to underestimate  
 316 tree density where there are ~25% less trees in the comparable ALS crown

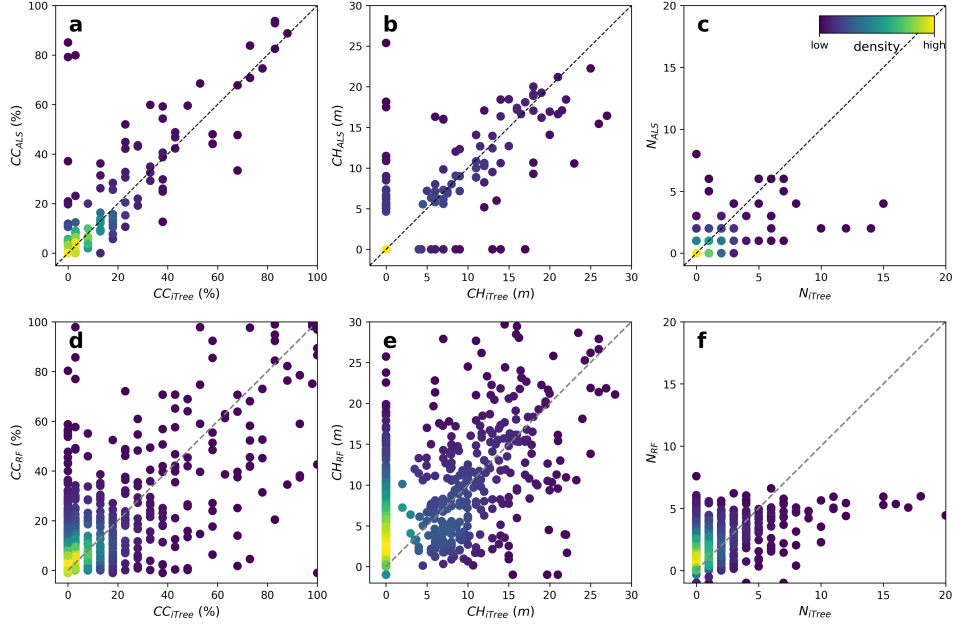


Figure 4: A comparison of iTree Eco with ALS (a-c,  $N=196$ ) and 20 m resolution Random Forest (d-f,  $N=741$ ) computed variables. Trees with a projected crown area  $\leq 10 \text{ m}^2$  were filtered from the iTree Eco database.

vector layer. This is particularly prevalent in plots with greater number of trees and indicates the inability of low resolution ALS to distinguish smaller or sub-dominant trees. However, as 90% of plots have  $\leq 4$  trees (equivalent to  $100 \text{ trees ha}^{-1}$ ), the inability of ALS to determine more dense plots will have a minimal impact on overall estimates.

Comparing Random Forest output to iTree Eco plot data highlights Random Forest's tendency to overestimate smaller values. This is particularly obvious when estimating canopy height where treeless plots are attributed with  $CH > 0 \text{ m}$  (Figure 4e). Further, tall trees with large crowns may overlap adjacent plots resulting in errors where  $CH_{iTree} = 0 \text{ m}$  and  $CH_{RF} \geq 10 \text{ m}$ .  $N_{RF}$  also underestimates densely forested plots, this is likely due to the underlying ALS derived training data.

iTree Eco data was also used to train a Random Forest (Table 4). A similar trend was seen where Random Forest slightly overestimates training data values (Figure 3); no systematic error correction was applied to iTree computed values owing to the small sample size. Mean  $CH_{iTree}$  and iTree

333 Eco  $CH_{RF}$  values are significantly lower than ALS trained estimates, this  
 334 is again likely due to the inability of ALS to detect sub-dominant trees.  
 335 Within sample standard deviation values are similar between iTree Eco and  
 336 ALS trained model output. However, standard deviation of mean values  
 337 (calculated from 25 iterations of Random Forest) is 3 times greater for pre-  
 338 dicted variables and an order of magnitude greater for  $\sum N$ ; this is likely  
 339 due to the smaller number of training plots.  $\sum N_{iTree}$  values are higher  
 340 than ALS estimated values, by  $\sim 2M$  trees when estimated over the Greater  
 341 London area. This would indicate a significant influence of positive outliers,  
 342 where the top 1% most densely treed plots contribute  $\sim 500K$  trees to the  
 343 overall estimate.

344 Maps of  $CC_{RF}$ ,  $CH_{RF}$  and  $N_{RF}$  for Greater London are presented in Fig-  
 345 ure 5. Spatial patterns of variable distributions are similar where higher val-  
 346 ues coincide with greenspace areas (compare with Figure 1). Areas towards  
 347 the edge of Greater London are home to the tallest and most dense trees,  
 348 particularly in the south-east and north-west of Greater London. Large  
 349 parks are also clearly visible e.g. Richmond Park in the south-west and  
 350 Hampstead Heath to the north. Differences in  $CH$  are evident between the  
 351 two resolutions owing to the method estimating the tallest tree a grid square;  
 352 however, the tallest trees are identified at both resolutions.

353 Model transferability was tested using two approaches; (1) estimation  
 354 of  $CC$  within Greater London and (2) applying local and Greater London  
 355 trained models to the UK city of Southampton. Within Greater London  
 356  $CC_{ALS}$  varied between 13.2% - 26.7% between “zones” (Table 5); this dis-  
 357 parity between training and test data sets can cause significant estimation  
 358 error. For example, using Zone 1 to train the model (a heavily forested  
 359 area on the outskirts of London, see Figure 1), Random Forest would over-  
 360 estimate  $CC_{ALS}$  by 2.5%. However, generally if the range in forest types is  
 361 adequately represented in the training data then error in mean  $CC_{RF} < 1\%$ .  
 362 RMSE of  $CC_{RF}$  increases by between 2% - 5%.

363 Applying Greater London trained models to Southampton produced  
 364 comparable results to that of locally trained models. Values for  $CC_{RF}$  for  
 365 locally and Greater London trained models were 16.4% (*std* 26.3%, *RMSE*  
 366 16.6%) and 15.8% (*std* 21.7%, *RMSE* 16.3%) respectively, Southampton’s  
 367 iTree Eco report estimated  $CC$  of 18.5% (reported as “tree cover”).  $CH$   
 368 values for locally and Greater London trained models were 7.5 m (*std* 8.0  
 369 m, *RMSE* 5.6 m) and 8.0 m (*std* 6.3 m, *RMSE* 5.7 m) respectively.  $\sum N$  was  
 370 overestimated by  $\sim 12\%$  using the Greater London model which estimated  
 371  $255,398 \pm 8,600$  trees compared to  $226,110 \pm 13,300$  trees using the locally  
 372 trained models; iTree Eco report estimated 267,500 trees (including trees

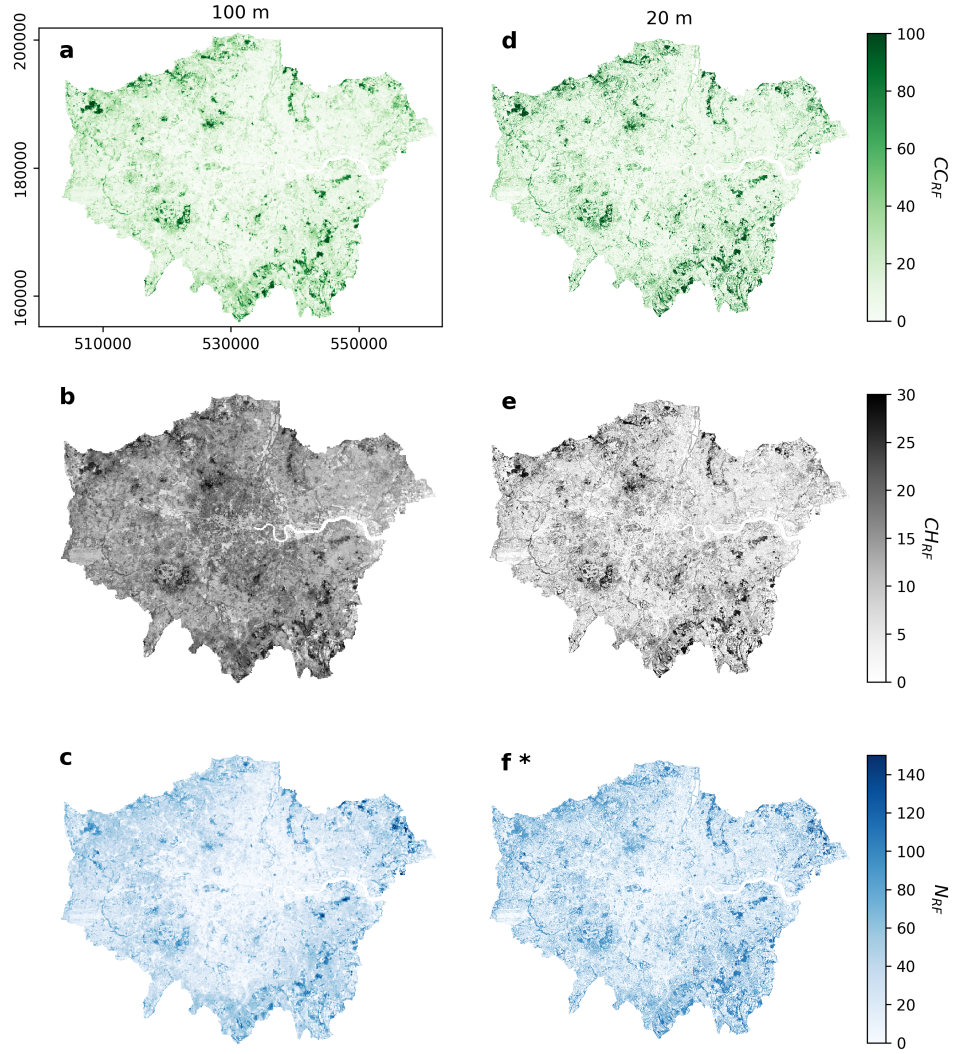


Figure 5: Maps of  $CC_{RF}$  (top),  $CH_{RF}$  (middle) and  $N_{RF}$  (bottom) at 100 m (left) and 20 m (right) resolutions. \* values are scaled to be equivalent of trees  $\text{ha}^{-1}$  so subplots c and f are comparable.

373 with a projected crown area  $\leq 10 \text{ m}^2$ ). These results would suggest that  
 374 urban forest structure can be inferred beyond the geographic domain of the  
 375 training data, albeit with the understanding that error is likely to increase.

Table 5: Random Forest estimated  $CC$  (%), computed at 20 m resolution, where the model has been applied to areas outside the geographic domain of the training data. Refer to Figure 1 for zones.

Training zone	<b>ALS<sub>train</sub></b>	<b>ALS<sub>test</sub></b>	<b>RF<sub>ALS</sub></b>			<b>RF+C<sub>ALS</sub></b>		
	mean	mean	mean	rmse	bias	mean	rmse	bias
1	26.7	18.1	23.1	19.6	5.0	20.6	21.5	2.5
2	13.9	18.6	18.6	20.0	0.0	18.8	22.3	0.2
3	13.2	18.8	18.6	19.1	-0.2	18.5	20.9	-0.3
4	21.6	17.4	20.9	18.2	3.5	19.6	20.4	2.2
5	18.6	19.0	23.5	18.7	4.5	20.9	19.8	1.9
6	15.5	19.1	20.8	19.2	1.7	19.7	20.0	0.6

#### 4. Discussion

Urban areas across the globe, currently home to ~10 billion trees [56], will play an increasingly important role in maintaining liveable cities in the face of population and climate pressures. Assessment tools are essential for maintaining and enhancing the ecosystem services provided by urban forests, as well as assessing the effectiveness of policy designed to improve this key asset. Presented here is a framework to use remote sensing methods to estimate wall-to-wall 3D forest structure in large and complex urban areas. New methods are required to deal with the particular difficulties of measuring urban compared to natural forest.

The framework uses freely available open-access data sets and techniques with the aim of creating a method that can be adopted for urban areas anywhere. Further, the framework is designed to be modular, as is demonstrated here by substituting ALS-derived estimates of forest structure with those from an iTree Eco survey. It is emphasised that this approach is meant to augment rather than replace existing protocols, such as iTree Eco; important properties captured by inventory methods (e.g. species composition, shrub metrics etc.) are currently beyond the capabilities of most remote sensing techniques.

The availability and suitability of training data are likely to be the most significant barriers to applying this method more generally. Here, ALS data produced a dense training data set of derived metrics (i.e. not directly measured) with a limited spatial extent, whereas the iTree data provide direct but very sparse measurements of canopy structure across the entire study area. Other sources of training data may include existing government tree databases or those produced from citizen science projects [57]

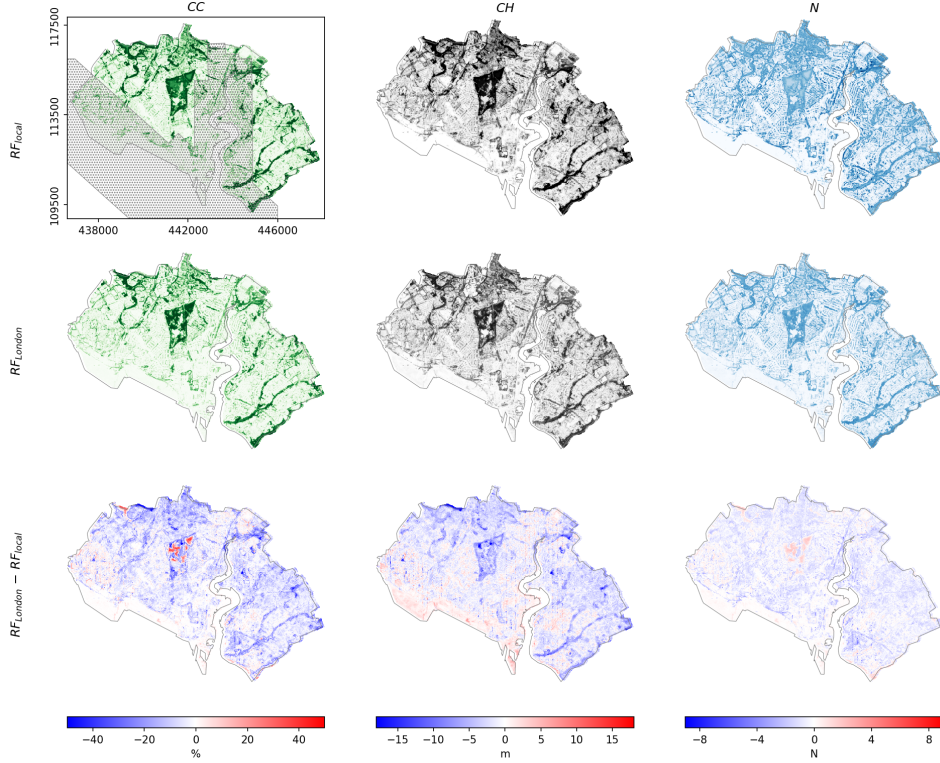


Figure 6: Canopy structure metrics derived for Southampton (See Figure 1). Forest structure estimated from; (top) a Random Forest model trained from local ALS (see ALS extent in top left); (middle) a Random Forest trained on data captured over London; and (bottom)  $RF_{London} - RF_{local}$ .

(e.g. <https://www.treezilla.org>) as well as other new technologies [58] (e.g. Google Street View [59] or terrestrial laser scanning [17]). Substituting data sets should be done with the knowledge of the expected changes in model uncertainty and variance.

There are a number of factors to consider when deciding upon predictor variables (e.g. availability, cost, pre-processing, spectral and spatial resolution etc.). For example, airborne imagery may provide high resolution data but with limited spectral information (e.g. RGB bands) and high acquisition costs, whereas satellite data are often multispectral and can be accessed free of charge but may be affected by cloud. Here, for both 100 m and 20 m analysis, the most important data sets were those derived from short wave infrared (SWIR) and “red edge” bands (Figure 7). This would indicate

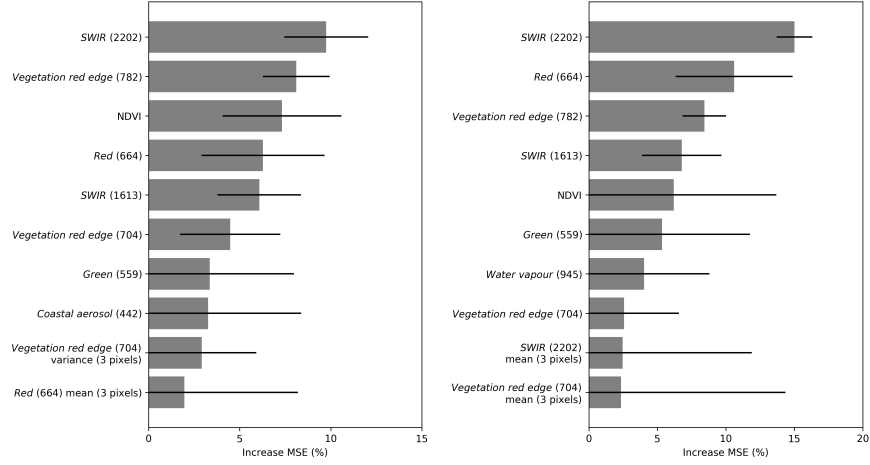


Figure 7: Variable importance for (left) 100 m analysis and (right) 20 m analysis. Italics are the names of Sentinel 2A bands and numbers in parentheses are wavelength in nm.

model predictions are driven by contrast in moisture content i.e. forest canopies withholding more moisture, and scattering of near-infrared radiation by leaves [26]. Further, this would suggest that this type of modelling is sensitive to change in forest structure (manifested in spectral reflectance) as opposed to being driven by abiotic factors (e.g. climate, topography etc.), and therefore temporal analysis would be possible. In contrast to natural forest systems [45], texture-based variables (e.g. image variance) are less important, highlighting the heterogeneity of the urban matrix. It should be remembered that the specific variables chosen here may be less suitable for other urban areas; therefore, a full suite of predictor variables should be considered before a more parsimonious set are chosen.

Incorporating synoptic data into tools such as iTree Eco may allow for a more robust estimate of the monetary benefits of urban forest i.e. capturing spatial heterogeneity. There are potential wider applications in other science domains beyond the inventory of urban forest, for example mapping habitat connectivity into and across urban areas (e.g. the "forest continuum" [60]), quantifying the impact of urban forest structure on public health or incorporating spatial explicit models of urban forest structure in air quality modelling. A "4D" analysis (where time is the fourth dimension) could potentially identify urban forest dynamics e.g. urban forest growth and removal. Suitable satellite imagery (e.g. Landsat) extends back to at least the mid-1980s and can be accessed through GEE. Contemporary training data could also be used, with the assumption that forest structure has remained

largely unchanged or by focusing on areas where this is true. As well as understanding past forest conditions the impact of new policy could be more effectively monitored.

## 5. Conclusions

Monitoring and assessment of our urban forests is crucial to maintaining viable, healthy and diverse forests as well as the ecosystem services city dwellers are reliant upon. This analysis demonstrates an open-access, flexible and modular framework that uses remote sensing data to assess urban forest structure (canopy height, canopy cover and tree density) across a large and heterogeneous urban forest. Results indicate that urban forest structure can be modelled accurately from remote sensing data sets and moreover produce wall-to-wall maps that have uses beyond the scope of just urban forest inventories. Further, these models are transferable to areas where appropriate training data does not exist. We suggest that methods such as these can be relatively easily and cheaply adopted by city planners, urban forest managers and greenspace advocates to ensure the future prosperity of urban forests globally.

## 6. Acknowledgements

PW and MD are funded by the NERC National Centre for Earth Observation (NCEO). MD acknowledges funding awards NE/N00373X/1, NE/P011780/1 and NE/K002554/1 as well as support by the EUHorizon2020 project (BACI project funded by the EU's Horizon 2020 Research and Innovation Programme under grant agreement 640176). The authors would like to thank Kieron Doick and Phil Handley (UK Forest Research) for providing access to the Greater London iTree Eco data set.



462 **Appendix A. A list of open-source software used.**

Table A.6: An alphabetical list of software used in the above methods. Dependencies are not included. \* indicates that package was accessed and installed using Conda, otherwise package was accessed through website. # some features may require a licence.

Package	Citation	Homepage	Version
GDAL*	[61]	<a href="https://gdal.org/">https://gdal.org/</a>	2.4.1
GeoPandas*	[62]	<a href="http://geopandas.org/">http://geopandas.org/</a>	0.0.5
LAStools#	[44]	<a href="https://rapidlasso.com/lastools/">https://rapidlasso.com/lastools/</a>	
QGIS	[63]	<a href="http://qgis.osgeo.org">http://qgis.osgeo.org</a>	3.4.7-Madeira
Pandas*	[64]	<a href="https://pandas.pydata.org/">https://pandas.pydata.org/</a>	0.24.2
Scikit Learn*	[52]	<a href="https://scikit-learn.org/">https://scikit-learn.org/</a>	0.20.3
Shapely*	[65]	<a href="https://pypi.org/project/Shapely/">https://pypi.org/project/Shapely/</a>	1.6.4.post2

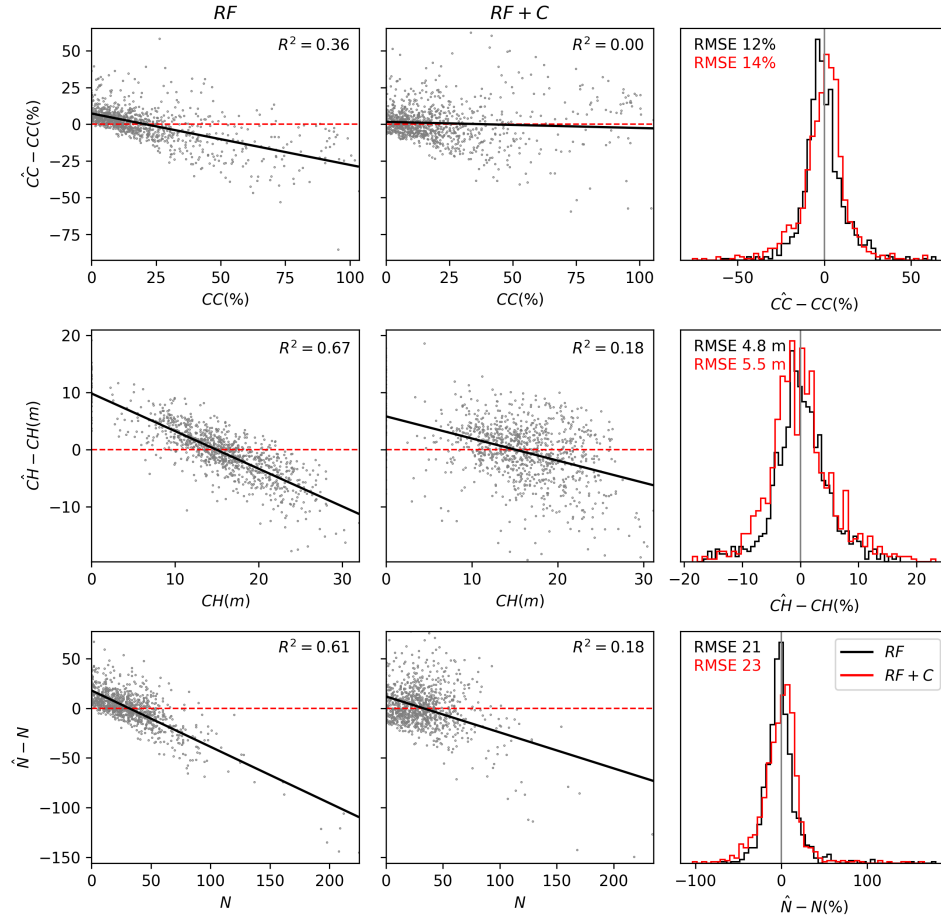


Figure B.8: Analysis of model residuals produced for analysis at 100 m for canopy cover  $CC$  (top row), canopy height  $CH$  (middle) and tree density  $N$  (bottom). (left) model residuals for untransformed output (middle) model residuals transformed using methods described in Section 2.3 and (right) histograms of residuals for untransformed and transformed data.

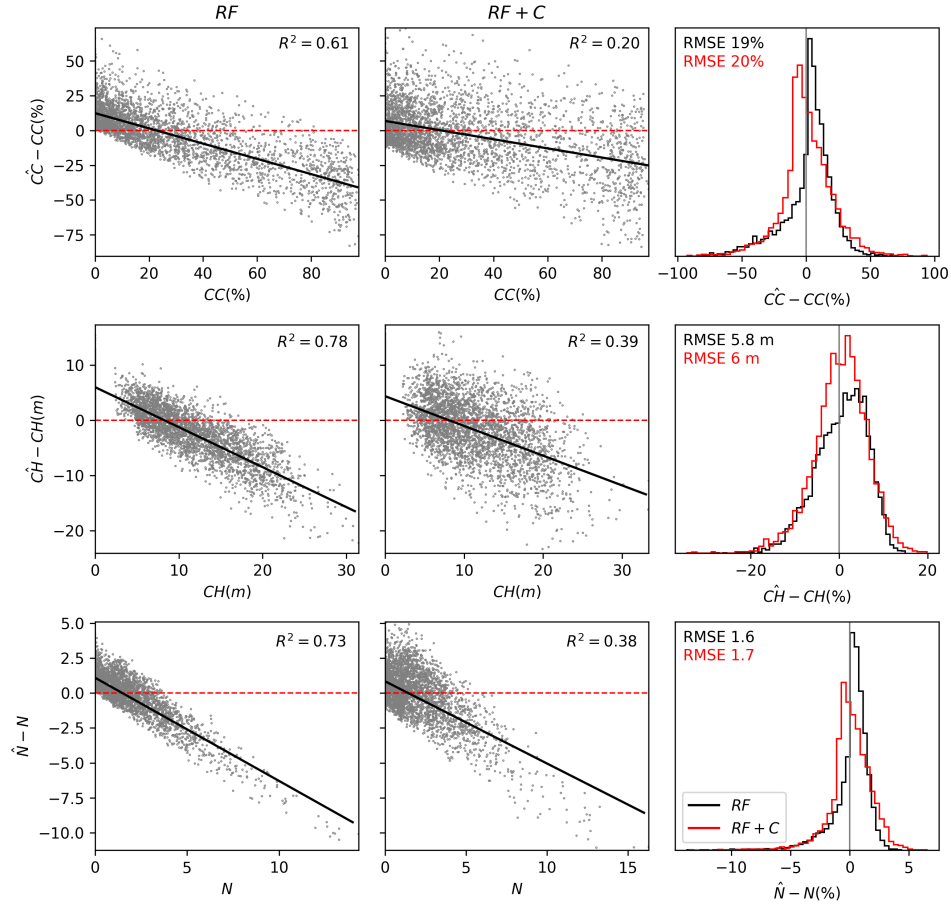


Figure B.9: Analysis of model residuals produced for analysis at 20 m. See Figure B.8 for notes.

## References

- [1] T. A. Endreny, R. Santagata, A. Perna, C. D. Stefano, R. Rallo, S. Ulgiati, Implementing and managing urban forests: A much needed conservation strategy to increase ecosystem services and urban wellbeing, Ecological Modelling 360 (2017) 328–335. URL: <http://dx.doi.org/10.1016/j.ecolmodel.2017.07.016><http://linkinghub.elsevier.com/retrieve/pii/S0304380017300960>. doi:10.1016/j.ecolmodel.2017.07.016.
- [2] H. Davies, K. J. Doick, P. Handley, L. O'Brien, J. Wilson, Delivery of

- ecosystem services by urban forests, Technical Report, Forestty Commission, Edinburgh, 2017.
- [3] D. J. Nowak, Atmospheric Carbon Dioxide Reduction by Chicago's Urban Forest, in: E. G. Mcpherson, D. J. Nowak, R. a. Rowntree (Eds.), Chicago's Urban Forest Ecosystem: Results of the Chicago Urban Forest Climate Project, USDA Forest Service General Technical Report NE-18, Radnor, PA, 1994, pp. 83–94.
- [4] K. J. Doick, H. J. Davies, J. Moss, R. Coventry, P. Handley, M. VazMonteiro, K. Rogers, P. Simpkin, The Canopy Cover of England's Towns and Cities: baselining and setting targets to improve human health and well-being, in: Trees, People and the Built Environment 3: Urban Trees Research Conference, Edinburgh, 2017. URL: [https://www.charteredforesters.org/wp-content/uploads/2019/01/Doick-et-al\\_Canopy-Cover-of-Englands-Towns-and-Cities\\_revised220317\\_combined.pdf](https://www.charteredforesters.org/wp-content/uploads/2019/01/Doick-et-al_Canopy-Cover-of-Englands-Towns-and-Cities_revised220317_combined.pdf).
- [5] Urban Forest Strategy: Making a great city greener 2012-2032, Technical Report, City of Melbourne, Melbourne, Australia, 2012. URL: [https://www.melbourne.vic.gov.au/Sustainability/UrbanForest/Documents/Urban\\_Forest\\_Strategy.pdf](https://www.melbourne.vic.gov.au/Sustainability/UrbanForest/Documents/Urban_Forest_Strategy.pdf).
- [6] AMEC Environment and Infrastructure, CITY OF VANCOUVER URBAN TREE CANOPY ASSESSMENT, Technical Report September, Department of Public Works, Victoria, British Columbia, 2011.
- [7] D. J. Nowak, D. E. Crane, J. C. Stevens, R. E. Hoehn, J. T. Walton, J. Bond, A ground-based method of assessing urban forest structure and ecosystem services, *Arboriculture and Urban Forestry* 34 (2008) 347–358.
- [8] USDA Forest Service, i-Tree Eco User's Manual v.6.0, Technical Report, 2017. URL: [http://www.itreetools.org/resources/manuals/Ecov6\\_ManualsGuides/Ecov6\\_UsersManual.pdf](http://www.itreetools.org/resources/manuals/Ecov6_ManualsGuides/Ecov6_UsersManual.pdf).
- [9] K. Rogers, K. Sacre, J. Goodenough, K. J. Doick, Valuing London's urban forest : results of the London i-Tree Eco Project., *Treeconomics*, London, 2015.
- [10] D. J. Nowak, R. Hoehn, D. Crane, J. Stevens, J. T. Walton, Assessing Urban Forest Effects and Values New York City's Urban Forest, Tech-

- 507 nical Report, USDA Forest Service, Northern Resource Bulletin, New-  
508 town Square, Pennsylvania, 2007. URL: <http://www.nrs.fs.fed.us>.
- 509 [11] L. Chaparro, J. Terrasdas, Ecological Services of Urban Forest in Bar-  
510 celona (2009) 103. URL: [https://www.itreetools.org/resources/](https://www.itreetools.org/resources/reports/BarcelonaEcosystemAnalysis.pdf)  
511 [reports/BarcelonaEcosystemAnalysis.pdf](https://www.itreetools.org/resources/reports/BarcelonaEcosystemAnalysis.pdf).
- 512 [12] R. Shojanoori, H. Z. M. Shafri, Review on the Use of Remote Sensing  
513 for Urban Forest Monitoring, *Arboriculture and Urban Forestry* 42  
514 (2011) 400–417.
- 515 [13] K. Roges, D. Hansford, T. Sunderland, A. Brunt, N. Coish, Meas-  
516 uring the ecosystem services of Torbay’s trees: th Torbay i-Tree Eco  
517 pilot project, in: M. Johnston, G. Percival (Eds.), *Trees, people*  
518 *and the built environment*, Forestry Commission, Edgbaston, Birm-  
519 ingham, UK, 2011, pp. 18–28. URL: [https://www.forestry.gov.uk/](https://www.forestry.gov.uk/pdf/Trees-people-and-the-buit-environment_Rogers.pdf/)  
520 [pdf/Trees-people-and-the-buit-environment\\_Rogers.pdf/](https://www.forestry.gov.uk/pdf/Trees-people-and-the-buit-environment_Rogers.pdf/).
- 521 [14] R. E. McRoberts, E. O. Tomppo, Remote sensing support for  
522 national forest inventories, *Remote Sensing of Environment* 110  
523 (2007) 412–419. URL: [http://linkinghub.elsevier.com/retrieve/](http://linkinghub.elsevier.com/retrieve/pii/S0034425707001721)  
524 [pii/S0034425707001721](http://linkinghub.elsevier.com/retrieve/pii/S0034425707001721). doi:10.1016/j.rse.2006.09.034.
- 525 [15] F. Barrett, R. E. McRoberts, E. Tomppo, E. Cienciala, L. T. Waser,  
526 A questionnaire-based review of the operational use of remotely sensed  
527 data by national forest inventories, *Remote Sensing of Environment*  
528 174 (2016) 279–289. doi:10.1016/j.rse.2015.08.029.
- 529 [16] Z. Zhu, Y. Zhou, K. C. Seto, E. C. Stokes, C. Deng, S. T. Pick-  
530 ett, H. Taubenböck, Understanding an urbanizing planet: Stra-  
531 tegic directions for remote sensing, *Remote Sensing of Environment*  
532 228 (2019) 164–182. URL: [https://doi.org/10.1016/j.rse.2019.](https://doi.org/10.1016/j.rse.2019.04.020)  
533 [04.020](https://doi.org/10.1016/j.rse.2019.04.020). doi:10.1016/j.rse.2019.04.020.
- 534 [17] P. Wilkes, M. Disney, M. B. Vicari, K. Calders, A. Burt, Estimating  
535 urban above ground biomass with multi-scale LiDAR, *Carbon Balance*  
536 *and Management* 13 (2018). doi:10.1186/s13021-018-0098-0.
- 537 [18] J. Tigges, T. Lakes, High resolution remote sensing for reducing uncer-  
538 tainties in urban forest carbon offset life cycle assessments, *Carbon Bal-*  
539 *ance and Management* 12 (2017). doi:10.1186/s13021-017-0085-x.

- [19] X. Li, W. Y. Chen, G. Sanesi, R. Laforzezza, Remote Sensing in Urban Forestry: Recent Applications and Future Directions, Remote Sensing 11 (2019) 1144. doi:10.3390/rs11101144.
- [20] J. T. Walton, D. J. Nowak, E. J. Greenfield, Assessing Urban Forest Canopy Cover Using Airborne or Satellite Imagery, Arboriculture & Urban Forestry 34 (2008) 334–340.
- [21] J. J. Richardson, L. M. Moskal, Uncertainty in urban forest canopy assessment: Lessons from Seattle, WA, USA, Urban Forestry & Urban Greening 13 (2014) 152–157. URL: <http://dx.doi.org/10.1016/j.ufug.2013.07.003><https://linkinghub.elsevier.com/retrieve/pii/S1618866713000800>. doi:10.1016/j.ufug.2013.07.003.
- [22] D. Ellingsworth, M. Binkley, S. Maco, i-Tree Canopy v6.1, 2019. URL: <https://canopy.itreetools.org/>.
- [23] Breadboard Lab, Measurement & spatial analysis of London’s tree canopy cover: 2018 methodology report, Technical Report, Greater London Authority, London, 2018.
- [24] Descartes Labs, Mapping All of the Trees with Machine Learning, 2018. URL: <https://medium.com/descarteslabs-team/descartes-labs-urban-trees-tree-canopy-mapping-3b6c85c5c9cc>.
- [25] M. Alonzo, B. Bookhagen, J. P. Mcfadden, A. Sun, D. A. Roberts, Mapping urban forest leaf area index with airborne lidar using penetration metrics and allometry, Remote Sensing of Environment 162 (2015) 141–153. doi:10.1016/j.rse.2015.02.025.
- [26] M. Alonzo, B. Bookhagen, D. A. Roberts, Urban tree species mapping using hyperspectral and lidar data fusion, Remote Sensing of Environment 148 (2014) 70–83. doi:10.1016/J.Rse.2014.03.018.
- [27] L. Liu, N. C. Coops, N. W. Aven, Y. Pang, Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data, Remote Sensing of Environment 200 (2017) 170–182. doi:10.1016/j.rse.2017.08.010.
- [28] S. M. Raciti, L. R. Hutya, J. D. Newell, Mapping carbon storage in urban trees with multi-source remote sensing data: Relationships between biomass, land use, and demographics in Boston neighborhoods,

- 574 The Science of the Total Environment 500-501 (2014) 72–83. doi:10.  
575 1016/j.scitotenv.2014.08.070.
- 576 [29] Food and Agricultural Organisation of the United Nations (FAO),  
577 Global Forest Resources Assessment 2020: Terms and Definitions, Tech-  
578 nical Report, Forest Resources Working Paper 188, Rome, 2018.
- 579 [30] R. Brennan, T. L. Webster, Object-oriented land cover classification of  
580 lidar-derived surfaces, Canadian Journal of Remote Sensing 32 (2006)  
581 162–172. doi:10.5589/m06-015.
- 582 [31] J. Schreyer, J. Tigges, T. Lakes, G. Churkina, Using Airborne LiDAR  
583 and QuickBird Data for Modelling Urban Tree Carbon Storage and  
584 Its Distribution—A Case Study of Berlin (2014) 10636–10655. doi:10.  
585 3390/rs61110636.
- 586 [32] M. A. Wulder, N. C. Coops, Satellites: Make Earth observations open  
587 access, Nature 513 (2014) 30–31. doi:10.1038/513030a.
- 588 [33] M. Whitworth, Laser surveys light up open data, 2015.  
589 URL: [https://environmentagency.blog.gov.uk/2015/09/18/  
590 laser-surveys-light-up-open-data/](https://environmentagency.blog.gov.uk/2015/09/18/laser-surveys-light-up-open-data/).
- 591 [34] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau,  
592 R. Moore, Google Earth Engine: Planetary-scale geospatial ana-  
593 lysis for everyone, Remote Sensing of Environment 202 (2017) 18–27.  
594 doi:10.1016/j.rse.2017.06.031.
- 595 [35] N. N. Patel, E. Angiuli, P. Gamba, A. Gaughan, G. Lisini, F. R.  
596 Stevens, A. J. Tatem, G. Trianni, International Journal of Applied  
597 Earth Observation and Geoinformation Multitemporal settlement and  
598 population mapping from Landsat using Google Earth Engine, Inter-  
599 national Journal of Applied Earth Observations and Geoinformation 35  
600 (2015) 199–208. doi:10.1016/j.jag.2014.09.005.
- 601 [36] K. Johansen, S. Phinn, M. Taylor, Mapping woody vegetation clearing  
602 in Queensland, Australia from Landsat imagery using the Google Earth  
603 Engine, Remote Sensing Applications: Society and Environment 1  
604 (2015) 36–49. doi:10.1016/j.rsase.2015.06.002.
- 605 [37] A. Rae, All Buildings in Great Britain, 2018. URL: [http://ajrae.  
606 staff.shef.ac.uk/buildings/](http://ajrae.staff.shef.ac.uk/buildings/).

- 607 [38] L. Breiman, Random forests, Machine learning 45 (2001) 5–32. URL:  
608 <http://link.springer.com/article/10.1023/A:1010933404324>.  
609 [arXiv:/dx.doi.org/10.1023%2FA%3A1010933404324](https://arxiv.org/abs/10.1023%2FA%3A1010933404324).
- 610 [39] Office For National Statistics (ONS) Population Estimates  
611 Unit, ONS Mid-Year Population Estimates - Custom Age  
612 Tables, 2017. URL: [https://data.london.gov.uk/dataset/  
613 ons-mid-year-population-estimates-custom-age-tables](https://data.london.gov.uk/dataset/ons-mid-year-population-estimates-custom-age-tables).
- 614 [40] Minisrty of Housing, Communities and Local Government, Land Use  
615 by Borough and Ward, 2005. URL: [https://data.london.gov.uk/  
616 dataset/land-use-ward](https://data.london.gov.uk/dataset/land-use-ward).
- 617 [41] R. L. Wilby, G. L. Perry, Climate change, biodiversity and  
618 the urban environment: a critical review based on London, UK,  
619 Progress in Physical Geography: Earth and Environment 30  
620 (2006) 73–98. URL: [http://journals.sagepub.com/doi/10.1191/  
621 0309133306pp470ra](http://journals.sagepub.com/doi/10.1191/0309133306pp470ra). doi:10.1191/0309133306pp470ra.
- 622 [42] London Tree Officers Association, LTOA Calculate London’s  
623 Tree Cover, 2010. URL: [https://www.ltoa.org.uk/news/  
624 159-calculate-londons-tree-cover](https://www.ltoa.org.uk/news/159-calculate-londons-tree-cover).
- 625 [43] Greater London Authority, London Authority Maintained  
626 Trees, 2016. URL: [https://data.london.gov.uk/dataset/  
627 local-authority-maintained-trees](https://data.london.gov.uk/dataset/local-authority-maintained-trees).
- 628 [44] M. Isenburg, LAsTools - efficient tools for LiDAR processing, 2015.  
629 URL: <https://rapidlasso.com/lastools/>.
- 630 [45] P. Wilkes, S. D. Jones, L. Suarez, A. Mellor, W. Woodgate, M. Soto-  
631 Berelov, A. Haywood, A. K. Skidmore, N. Baghdadi, P. S. Thenkabail,  
632 Mapping Forest Canopy Height Across Large Areas by Upscaling ALS  
633 Estimates with Freely Available Satellite Data, Remote Sens. 7 (2015)  
634 1–25. doi:10.3390/rs70x000x.
- 635 [46] A. Mellor, A. Haywood, C. Stone, S. D. Jones, The performance  
636 of random forests in an operational setting for large area sclerophyll  
637 forest classification, Remote Sensing 5 (2013) 2838–2856. doi:10.3390/  
638 rs5062838.
- 639 [47] Y. Du, Y. Zhang, F. Ling, Q. Wang, W. Li, X. Li, Water Bodies’  
640 Mapping from Sentinel-2 Imagery with Modified Normalized Difference



641 Water Index at 10-m Spatial Resolution Produced by Sharpening the  
642 SWIR Band, *Remote Sensing* 8 (2016) 354. doi:10.3390/rs8040354.

643 [48] L. Breiman, Bagging predictors, *Machine Learning* 24 (1996) 123–140.  
644 doi:10.1007/BF00058655.

645 [49] M. Simard, N. Pinto, J. B. Fisher, A. Baccini, Mapping forest canopy  
646 height globally with spaceborne lidar, *Journal of Geophysical Research*  
647 116 (2011) 1–12. doi:10.1029/2011JG001708.

648 [50] O. S. Ahmed, S. E. Franklin, M. A. Wulder, J. C. White, Characterizing  
649 stand-level forest canopy cover and height using Landsat time series,  
650 samples of airborne LiDAR, and the Random Forest algorithm, *ISPRS*  
651 *Journal of Photogrammetry and Remote Sensing* 101 (2015) 89–101.  
652 doi:10.1016/j.isprsjprs.2014.11.007.

653 [51] A. Baccini, N. T. Laporte, S. J. Goetz, M. Sun, H. Dong, A first  
654 map of tropical Africa’s above-ground biomass derived from satellite  
655 imagery, *Environmental Research Letters* 3 (2008) 1–9. doi:10.1088/  
656 1748-9326/3/4/045011.

657 [52] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion,  
658 O. Grisel, M. Blondel, G. Louppe, P. Prettenhofer, R. Weiss, V. Du-  
659 bourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Per-  
660 rot, E. Duchesnay, *Scikit-learn: Machine Learning in Python* 12 (2012)  
661 2825–2830. doi:10.1007/s13398-014-0173-7.2.

662 [53] G. Zhang, Y. Lu, Bias-corrected random forests in regression, *Journal*  
663 *of Applied Statistics* 39 (2012) 151–160. doi:10.1080/02664763.2011.  
664 578621.

665 [54] M. Vaz Monteiro, K. J. Doick, P. Handley, Allometric relationships for  
666 urban trees in Great Britain, *Urban Forestry & Urban Greening* 19  
667 (2016) 223–236. doi:10.1016/j.ufug.2016.07.009.

668 [55] E. Mutch, K. Doick, H. Davies, P. Handley, M. Hudson, S. Kiss, L. Mc-  
669 Culloch, K. Parks, K. Rogers, K. Schreckenberg, Understanding the  
670 value of Southampton’s urban trees. Results of the 2016 i-Tree Eco  
671 survey, Technical Report, University of Southampton, Forest Research,  
672 Treeconomics and Southampton City Council, Southampton, 2017.

673 [56] T. A. Endreny, Strategically growing the urban forest will  
674 improve our world, *Nature Communications* 9 (2018) 1160.

- 675 URL: <http://www.nature.com/articles/s41467-018-03622-0>.  
676 doi:10.1038/s41467-018-03622-0.
- 677 [57] L. A. Roman, B. C. Scharenbroch, J. P. Östberg, L. S. Mueller, J. G.  
678 Henning, A. K. Koeser, J. R. Sanders, D. R. Betz, R. C. Jordan,  
679 Data quality in citizen science urban tree inventories, *Urban Forestry*  
680 and *Urban Greening* 22 (2017) 124–135. doi:10.1016/j.ufug.2017.  
681 02.001.
- 682 [58] S. A. Nitoslowski, N. J. Galle, C. K. van den Bosc, J. W. Steenberg,  
683 Smarter ecosystems for smarter cities? A review of trends, technologies,  
684 and turning points for smart urban forestry, *Sustainable Cities and*  
685 *Society* 51 (2019) 101770. doi:10.1016/j.scs.2019.101770.
- 686 [59] X. Li, C. Zhang, W. Li, R. Ricard, Q. Meng, W. Zhang, Assessing  
687 street-level urban greenery using Google Street View and a modified  
688 green view index, *Urban Forestry & Urban Greening* 14 (2015) 675–  
689 685. doi:10.1016/j.ufug.2015.06.006.
- 690 [60] J. Pirnat, D. Hladnik, The concept of landscape structure, forest con-  
691 tinuum and connectivity as a support in urban forest management and  
692 landscape planning, *Forests* 9 (2018). doi:10.3390/f9100584.
- 693 [61] GDAL/OGR contributors, {GDAL/OGR} Geospatial Data Abstrac-  
694 tion software Library, 2019. URL: <https://gdal.org>.
- 695 [62] K. Jordahl, J. V. d. Bossche, J. Wasserman, J. McBride, J. Ger-  
696 ard, J. Tratner, M. Perry, C. Farmer, *geopandas*, 2019. doi:10.5281/  
697 *zenodo.2705946*.
- 698 [63] QGIS Development Team, QGIS Geographic Information System, 2019.  
699 URL: <http://qgis.osgeo.org>.
- 700 [64] W. McKinney, Data Structures for Statistical Computing in Python,  
701 in: S. van der Walt, J. Millman (Eds.), *Proceedings of the 9th Python*  
702 *in Science Conference*, 2010, pp. 51–56.
- 703 [65] S. Gillies, others, *Shapely: manipulation and analysis of geometric ob-*  
704 *jects*, 2019. URL: <https://github.com/Toblerity/Shapely>.