# Quantifying urban forest structure in Greater Manchester with open-access remote sensing datasets

by

Philip Home

A dissertation submitted in partial fulfillment
of the requirements for the degree of
MRes Geospatial Data Science
Faculty of Engineering, University of Nottingham.

Nottingham Geospatial Institute, School of Engineering Surveying and Space Geodesy, 30 Triumph Road, Lenton, Nottingham, NG7 2TU, United Kingdom.

Date: Friday 26th August 2021

**Student Number:** 20385861

**Declaration:** "I hereby certify that this work is my own, except where otherwise acknowledged, and that it has not been submitted previously for a degree at this, or any other university."

**Supervisor:** Professor Doreen Boyd

## Acknowledgements

This research is funded by the EPSRC as part of a studentship in the Geospatial Systems CDT. Professor Doreen Boyd, Oliver Baines and Bryan Cosgrove with their assistance thus far in the project.

## Table of Contents

## List of Tables

## List of Figures

## Abstract

A growing body of evidence links the adverse impacts of expanding urbanism including increased air pollution, and exposure to heat stress with the removal of vegetation within cities. As the global population is estimated to reach 10 billion by 2050, urban trees and extended green infrastructure are advocated as a remedy to the effects of increasing urbanisation through delivering a multitude of ecosystem services including pollution abatement, reduction of urban heat islands and social benefits. To accurately quantify the services afforded by urban forests, it is vital to measure the extent and structure of urban forests, including through time, in addition for assessing the success of policy to maintain and promote green infrastructure assets. Current ground fieldwork methods rely on plot networks to measure a range of metrics across the tree population; these methods are locally comprehensive however do not fully describe the spatial heterogeneity of the urban fabric, given the limited sampling and often laborious data collection. The increasing availability and access to remote sensing/earth observation datasets provide an opportunity to collate synoptic measurements across large regions. Direct measurements though active sensors, particularly LiDAR, have seen wide adoption when measuring forest structure, however surveys can be expensive, and coverage limited. Fusing LiDAR with satellite imagery though machine learning methods such as Random Forests can drastically increase coverage through capturing complex non linear relationships. A framework is presented to estimate forest structure using open access data and software across Greater Manchester. This workflow estimates three forest structure metrics, canopy cover, canopy height and tree number/density. Random forest models were trained with airborne Environment Agency LiDAR, and predictor variables derived from Sentinel 2 and ancillary climatic and topographic datasets. Results indicate estimates in 2018, mean canopy cover of 14.9% (RMSE = 13.75), mean canopy height of 14.83m (RMSE = 6.14m) and home to ~2.6 million trees (RMSE = 0.95 per pixel). Results appear to illustrate higher canopy cover than i-Tree ground data but lower tree density and canopy heights. Altering input resolution was found to change structure estimations, attributed to methodological issues. Forest structure estimates were found to change from 2018 to 2021 indicating net decreases in canopy cover and number of trees, while average canopy height was found to increase, although change distribution of metrics

across boroughs is not equal. Presented methods can augment traditional inventory methods and can assist urban forest/land managers to produce consistent monitoring information to support the sustainability of urban forests worldwide.

## 1. Introduction

Globally, there exists a shared experience of increasing urbanisation and growth of cities (Nesbitt et al., 2017). The United Nations (UN) estimates the percentage of global population living in urban areas will increase from around 56% today to 68% by 2050, around 6.7 billion people (UN Department of Economic and Social Affairs, 2018). While cities are considered factories of opportunities for employment and education, left unmanaged can be responsible for considerable environmental degradation; already contributing to 70% of global greenhouse gases (UN HABITAT, 2011), this is only expected to increase as the proportional urban population grows. As a function of increased urban population, resource demand exacerbates several social, environmental, and economic issues in areas of already high vulnerability (Walters & Sinnett, 2021).

Challenges posed by increasing urbanism are varied and complex, including gas regulation/air filtering, micro-climate regulation, rainwater drainage, noise reduction (disturbance regulation), sewage treatment, among others (Bolund & Hunhammar, 1999). The severity of urban impacts are such that the United Nations (UN) in 2015 developed the Sustainable Development Goals (SDGs), the Food and Agriculture Organisation (FAO) mapped how urban forests advance nine SDGs (FAO, 2016).While SDG 11 definitively specifies ambitions for "Sustainable and Resilient Cities and Human Settlements" given the rise of urban populations cities are generally considered to be crucial to the entire SDG agenda (UN - HABITAT, n.d.).

Cities are comprised of built components including buildings, roads, bridges and other anthropogenic structural amenities; known as grey infrastructure (Cameron & Blanuša, 2016a). As an antonym to grey infrastructure, the term green infrastructure (GI) was coined. Green infrastructure according to Natural England (2009) is defined as "A strategically planned and delivered network comprising the broadest range of high quality green spaces and other environmental features….delivering ecological services and quality of life benefits required by the communities it serves". GI is composed of a variety of green landscape typologies including parks, nature reserves, gardens, river corridors, etc. Urban trees/forests are a component of green infrastructure, as such are a critical component within the urban fabric to promote sustainability and deliver ecosystem services to communities. As the urban

matrix encroaches into rural areas to accommodate for growing populations, nurturing urban green infrastructure is of rising importance for quality of life benefits for residents.

Doick et al. (2016) defines urban forests as 'all the trees in the urban realm – in public and private spaces, along linear routes and waterways and in amenity areas. It contributes to green infrastructure and the wider urban ecosystems". This definition also encompassed urban forests, both biotic and abiotic components. Urban forests are a nature based solution, natural methods to leverage ecosystems to effectively address or ameliorate societal challenges. The ecological community has long known urban trees afford society ecosystem services including benefits carbon sequestration, pollution abatement, mitigation of the urban heat island and flooding (Monteiro et al., 2020; Seddon, 2022). Additionally, further benefits through can be gained via enhancement of biodiversity, physical and mental wellbeing, and promotion of community cohesion (Roy et al., 2012; Salmond et al., 2016).

To that end governments are setting targets to increase urban tree cover at multiple scales (Doick et al, n.d.). The city of Bristol intends to double canopy cover by 2045 (Walters & Sinnett, 2021), Greater Manchester's (GM) ambitious target of planting three million trees by 2035 is intended to reap these benefits and improve access to green space (CITY OF TREES, n.d.; GMCA, 2019), a key indicator of aggregated urban population health according to the World Health Organisation (PUBLIC HEALTH ENGLAND, 2020; Salmond et al., 2016; Huang et al., 2017). To fully understand magnitude of benefits associated with urban forest, information regarding forest/tree structure is required to provide the basis for natural capital estimates (Nowak et al., 2008).

To assess the success of urban forest policy measures an in-depth understanding of current forest inventory is required (Chrysoulakis et al., 2021; Baines et al., 2020). Systematic frameworks for assessing urban forest inventory such as the i-Tree Eco protocol have been established by the United States Forest Service (Raum et al., 2019), an adaptation of the Urban Forest Effects model (UFORE). This methodology has seen ubiquitous take up across the UK and US (Monteiro et al., 2020; Lin et al., 2021). This methodology standardizes field techniques including defining features of individual trees such as diameter breast height (DBH), species, cover and using fixed size sampling plots. Once a representative sample is taken, the i-Tree Eco method extrapolates across larger regions to estimate economic value for the regions ecosystem services, useful for urban planners and land managers to justify GI.

A 2018 i-Tree Eco project undertaken in GM estimated 11,321,386 trees, generating an annual economic value of £33.3 million from associated ecosystem services with an estimated replacement figure of £4.7 billion (CITY OF TREES, n.d.). Approximately 15.7% of GM is estimated to be under canopy, around the national average and above London (Doick et al, n.d.). Although these results are statistically sound, they are based upon extrapolated ground data from many ground samples and have the potential to miss important data points. Other estimates from the 2010 Greater Manchester Tree Audit mapped individual trees and woodland, determined in similar canopy cover results of 16% across GM, indicating a small decrease in an eight year period (CITY OF TREES, n.d.), while Bluesky's National Tree Map estimates nearly 4.8 million trees in the same year (BLUESKY, 2017). To fully grasp the impact of the urban forests within the urban fabric, comprehensive estimates are required spanning the entirety of the specified region, given impacts will vary dependent on location (Donovan, 2017). Larger benefits may arise from specific areas and therefore influence optimal decision making for further tree planting and retention. Given GM is an aggregate of ten separate boroughs, higher resolution estimates may be of larger importance for improved borough specific management strategies, to assess individual borough inventory and also evaluate GI policy on a smaller scale.

In situ fieldwork is often laborious, time consuming and potentially resource expensive even if locally comprehensive (Baines et al., 2020). Remote sensing (RS) through satellite imagery offers a different approach through synoptic monitoring options operationally spanning vast geographic regions, with high spatial and temporal resolutions (Baines et al., 2020; Wilkes et al., 2018). Given the Flexibility and utility of RS, utilizing several sensors hosted on different platforms, RS has been applied to many different use cases (Ruiz et al., 2017). RS sports data collected from several sources including passive multi and hyperspectral sensors and active sensors such as RADAR (Radio Detection and Ranging) and LiDAR (Light Detection and Ranging). LiDAR in particular has seen wide uptake in ecological studies concerning 3D vegetation structure; including height, canopy cover (Wilkes et al., 2018; Ahmed et al., 2015; Alonzo et al., 2016; Bruggisser et al., 2019; Neuville et al., 2021) and also species mapping (Alonzo et al., 2014) including in urban environments. Active and passive sensing techniques already been widely adopted for building forest inventories (Puissant et al., 2014; Hanssen et al., 2021; Matasci et al., 2018).

LiDAR sensors have been mounted on different platforms including spaceborne, airborne and terrestrial/mobile, with different suitability at varying scales (Lechner et al., 2020). Terrestrial

LiDAR Surveys (TLS) and Mobile LiDAR Surveys (MLS) tend to be applied to plot level metrics (Bruggisser et al., 2019), further the "bottom up" approach of MLS and TLS provide detailed information on how humans view forests, and the rapid deployment of TLS/MLS could provide time sensitive insights into conservation and management. Airborne LiDAR Surveys (ALS) are mainly applied to regional/landscape scale forestry studies (Baines et al., 2020; Ahmed et al., 2015; Roussel & Auty, 2021; Bruggisser et al., 2019). ALS are undoubtedly more useful over larger regions than TLS however insights can be obscured by dominant canopy and the largest trees (Bruggisser et al., 2019). However, ALS has the benefit of flexibility, being able to be tasked for response to specific events such as treefall (Lechner et al., 2020), additionally resolution tends to be higher than spaceborne datasets. Spaceborne sensors such as the Global Ecosystem Dynamics Investigation (GEDI) hosted on the International Space Station are applied where ALS would be too costly or impractical such as global level forest structure metrics (Potapov et al., 2021; Hansen et al., 2013).

Many studies have been undertaken combining LiDAR and information from other sources and sensors to estimate and model forests metrics(Alonzo et al., 2014; Wilkes et al., 2018; Alonzo et al., 2016; Neuville et al., 2021; Baines et al., 2020). Through combining with other information sources, some of the limitations of LiDAR such as limited spatial extent can be overcome. However, challenges still remain when applying LiDAR for the purpose of urban forest inventory and structure mapping. Firstly, the highly complex nature of the urban matrix encompassing features such as energy infrastructure, roads, buildings, rivers, trees and other ground features, creates difficulty in the LiDAR filtering process (Zhang et al., 2015). Second, the spatial heterogeneity of urban forests is a hurdle, unlike natural forests, urban forests generally do not share similar canopy profiles, differing heights and crown widths and can often be isolated or single trees (Zhang et al., 2015; Baines et al., 2020). Urban forests may also be a rich mixture of broadleaf and conifer species, which blend into each other during LiDAR surveys (Zhang et al., 2015).

This study applies freely available satellite imagery and ALS datasets, in a Random Forest (RF) machine learning method, for forest structure estimation and distribution in GM. Three forest structure parameters are chosen: Canopy Height (CH), Canopy Cover (CC) and number of trees present (N). CH relates to the maturity/age of the tree and can be used allometric equations in aboveground biomass (AGB) estimations (Wilkes et al., 2018). Canopy Cover is a useful metric to determine urban forest extent. Number of trees helps to indicate the importance of each individual tree and allows for comparison between urban forests.

Research outputs are spatially explicit maps of 3D forest structure metrics at 20m/100m resolution across the entirety of GM. The method described in this study is designed to be entirely open source, with the aim to adhere to FAIR (Findable, Accessible, Interoperable and Reusable) principles as championed by UKRI. As such it is hoped this method will be applied to other heterogeneous urban spaces in regions, particularly regions with fewer available resources such as in the global south. The created maps will allow future works to study the relationship between urban forests and their impacts on the urban fabric at finer resolutions than previously and assist land managers to monitor GI targets.

## 1.1 Aims and Objectives

To present a framework to produce Greater Manchester wide estimates of urban forest structure metrics: number of trees (N), canopy height (CH) and canopy cover (CC). To be realised through applying open source data within a Random Forest model.

- Examine model performance and spatial distribution of estimated urban forest metrics
- Comparison of estimated metrics with those of the GM i-Tree Eco report
- Examine changes in forest structure metrics from 2018 to 2021

# 2. Literature Review

## 2.1 Ecosystem services from urban forests

Urban forests offer an array of ecosystem services, these services are generally considered in physical expressions, relating to the effect on the local environment. (Bolund & Hunhammar, 1999) define ecosystem services as benefits derived from the ecosystems which improve human wellbeing. This definition has been later adapted by Boyd & Banzhaf (2007) and Fisher & Kerry Turner (2008), creating a dialogue to delineate between semantics such as "ends" and" means"; "services" and "benefits". The concept of ecosystem services still remains somewhat ambiguous to different stakeholders, confusion among the terms ecosystem functions, services, and disservices makes consistent measurement and valuation difficult, somewhat limiting their usefulness and relevance if a societal goal is to manage urban forests for societal advantage (Escobedo et al., 2011). Nevertheless, consensus exists in that urban trees impact the urban matrix through ecosystem services/benefits.

Utilising the i-Tree Eco methodology set out by the USDA Forestry Service, a suite of structural metrics have been produced for cities across the UK. Coordinated by Forest Research; the UK's principal organisation for forestry related research, results of which are summarised in Table 1. Results of several other i-Tree Eco reports are due to be published including reports for Derby, Vale of Glamorgan and Cambridge. Table 2 indicates the high combined ecosystem service monetary value for urban forests, illustrating the cost if left unmanaged.

*Table 1:i-Tree Eco survey results from UK studies*

| *Location* | *Number of Trees* | *Canopy Cover* | *Annual Natural Capital Value* | *Survey Year* | *Source* |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| *London* | 8,421,000 | 14% | £132.7m | 2014 | (Forest Research, 2015b) |
| *Newport* | 259,900 | 12% | £143,000 | 2019 | (Forest Research, 2020) |
| *Petersfield* | 60,570 | 15.1% | £75,000 | 2016 | (Forest Research, 2017b) |
| *Edinburgh* | 712,000 | 17% | £1.82m | 2011 | (Forest Research, 2017a) |
| *Glasgow* | ~2,000,000 | 15% | £4.5m | 2013 | (Forest Research, 2015a) |
| *Southampton* | 267,000 | 18.5% | £1.29m | 2016 | (Forest Research, 2017c) |

On both a local and global scale, one ecosystem service provided by urban trees reducing the impact of urbanisation is carbon dioxide sequestration. Urban forests are well placed to sequester carbon due to the proximity to direct emitters i.e., vehicle emissions (Wilkes et al., 2018). Absorbed by plant tissue via photosynthesis and subsequently stored in woody tissue as biomass; this sequestered carbon has a significant monetary impact to the surrounding area. As such biogenic carbon sequestration by urban trees is estimated at 2.36 M tonnes and has been allocated an evaluation of £4.5M value per annum in Greater London (Forest Research, 2015b). This London study was applied to other global megacities including Beijing, Los Angeles and Mexico City to estimate the annual benefits of forest in these urban environments, revealing sizable median annual values of $505 million and an additional $7.9 billion in total value of carbon storage (Endreny et al., 2017). Despite this seemingly large

contribution mitigating carbon emissions, urban forests are generally not considered in the global carbon cycle, seemingly owing to the proportionally small spatial area as part of global forest contribution. Consideration of urban forest carbon cycling is increasingly important given the urban land expansion is one of the most visible irreversible and rapid land cover transformations in contemporary human history with urban land projections of up to 3.6 million $km^2$ by 2100 (Gao & O'Neill, 2020) already encompassing an estimated 10 billion urban trees (Endreny, 2018). Therefore, tools to monitor stored carbon in urban green infrastructure are crucial to ascertain for accurate global assessments.

Conversely, urban trees can also be a nuisance to society, also termed "disservices", in some situations (Roy et al., 2012). Increased costs involved in management of urban forests, through maintenance, planting/establishment and irrigation (Escobedo et al., 2011), and potential false benefit perception of trees in all urban scenarios, while in reality urban forests may not always be appropriate or beneficial to all stakeholders. One of the most discussed disservices is the emission of biogenic volatile compounds, that can lead to secondary formation of ground level ozone which contributes towards respiratory illnesses (Roy et al., 2012). Urban trees have the potential to damage grey infrastructure, where rapidly grown species can damage pavements and roads through root expansion increasing urban maintenance costs (Escobedo et al., 2011). Disservices can also directly affect human populations, through increased exposure to wild animals, safety issues with tree litter fall, but perhaps more importantly population health challenges such as increased allergenic response and refugia for vector-spread diseases (Escobedo et al., 2011; Doick et al., 2017; Cameron & Blanuša, 2016b). Benefits that arise from urban trees can lead to adverse impacts dependent on spatial and environmental context, for example during summer periods urban tree shading effects can be beneficial, interception of solar radiation leads to mitigation of the urban heat island effect (UHI), while the same effect during winter seasons can result in higher heating costs incurred to residents (Doick et al., 2017). Similarly, air purification through assimilation is beneficial to overall city air pollution abatement, however the street canyon effect of densely planted trees reduces wind speed to the extent where pollutants are trapped below canopy reducing air quality and posing increased risk to pedestrians (Doick et al., 2017; Escobedo et al., 2011).

Nevertheless, literature suggests the social benefits provided by urban forests outweigh the environmental and economic costs of maintaining them (Escobedo et al., 2011) and disservices can be mitigated through location and species management (Doick et al., 2017;

Donovan, 2017). Service and disservice interaction appears highly complex, and such is the focus of several research groups. Given the influence of location on service delivery, regular monitoring and inventories of urban trees are required to optimise ecosystem service delivery while mitigating disservices. Fieldwork can be time consuming, resource intensive and spatially coarse, lacking full city coverage. Developments in remote sensing in addition to advances in computing and cloud computing have aided and improved forest structure monitoring (Donager et al., 2021).

## 2.2 Remote Sensing of urban forest structure

Remote sensing of urban forests remains difficult due to the inherent challenges associated with its structure, particularly the spatially heterogeneous and fragmented nature of urban ecosystems (Zhu et al., 2019). In spite of difficulties associated with urban forest remote sensing numerous successful studies assessing urban forest composition and structure have been completed in recent years (Baines et al., 2020; Li et al., 2019).Literature suggests there is emphasis in forest structure mapping studies to focus on canopy cover and canopy height. BLUESKY (2017) were able to produce England, Wales and Scotland wide estimates for individual trees (crown polygons) and heights using passive sensing methods. Aerial photography and photogrammetric techniques were utilized to produce accuracies of over 90% in canopy cover and +/- 1.5m height accuracy. Hassaan et al. (2016) was also successful in implementation of tree defining algorithm using optical imagery to a 0.72 accuracy. The iTree Eco Canopy tool also uses optical imagery from Google Maps to allow users to assess if a point is located under canopy ('i-Tree Eco | i-Tree', 2022).

Baines et al., (2020) created Greater London canopy cover and canopy height estimates at both 20m and 100m resolution applying Environment Agency airborne LiDAR surveys and Sentinel 2 imagery. Over a much larger scale, Potapov et al., (2021) again combining LiDAR, this time from the Global Ecosystem Dynamics Investigation instrument aboard the International Space Station with analysis ready satellite imagery from the LandSat instrument to produce global canopy height estimates at a 30m spatial resolution for the year 2019. Both studies combine LiDAR and medium resolution satellite imagery. Variables used to calibrate the models in these studies are mainly surface reflectance inputs however also incorporated are variables that could explain the spatial distribution of forest structure such as topography and precipitation (Wilkes et al., 2015).

The LandSat satellite programme is a popular data source in forest structure studies,(Potapov et al., 2021; Ahmed et al., 2015). The series provide 30m resolution over an extended period, additionally the series is freely available, as the cost of commercial imagery may be prohibitive.

Research assessing forest structure have exploited multispectral imagery to determine bands that identify vegetation, such as the Short Wave Infrared (SWIR) and "red edge" bands (Baines et al., 2020), indicating models are driven by the contrast across water moisture content as forests scatter more near-infrared radiation by tree leaves due to the withholding of moisture by canopy (Alonzo et al., 2014, 2016). To further delineate vegetated surface, the combination of several bands into explicit indices are vital to detect specific properties of a surface such as greenness and moisture content.

Spectral indices (SIs) are a key part to many forest mapping studies by manipulating bands to create explicit contrast. In both landcover and forest structure mapping, a ubiquitously popular index is the Normalised Difference Vegetation Index (NDVI) (Ahmed et al., 2015), calculated through manipulation of the red (visible) and Near Infrared (NIR) bands. Vegetated surfaces indicate a high NDVI value as internal plant cell reflections lead to strong NIR reflectance and strong absorption in the visible red wavelengths due to the presence of chlorophyll. NDVI has been utilized in studies regarding height and extent (Ahmed et al., 2015; Potapov et al., 2021), species classification (Alonzo et al., 2014) and land cover classification (Huang et al., 2020).

Reliable and accurate change detection is a key pillar of RS, longitudinal datasets such as the LandSat archive are particularly suited to temporal analysis to indicate the success of multi-decadal policies (Ahmed et al., 2015). Features with spectral similarity may be distinguished through including a temporal aspect, in fact temporal aggregation of data maybe crucial to represent the phenological variation across time (Huang et al., 2020), shown to facilitate forest height modeling (Potapov et al., 2021). Ahmed et al. (2015) utilized a temporal approach in forest structure mapping, using forest disturbance history as a metric to better distinguish forest stands between mature and young forests, leading to improvements in RMSE of ~20%. Further, both (Mellor et al., 2013) and (Wilkes et al., 2015) both apply temporal texture to represent phenological variance in structure studies.

"Texture" referring to the change in image intensity gradients, exploring the heterogeneity of spectral magnitude throughout an image. Accounting for texture and not only the absolute magnitude is vital part to many structure studies. By calculating statistics such as averages and variance within windows, a measure of variation can be ascertained. (Ruiz et al., 2017) notes the combination of texture metrics in combination with spatial metrics provided a six percentage better distinction accuracy in land use classification than using spatial bands in isolation. Further, in vegetation height mapping Lang et al. (2019), found the benefit of using texture metrics when using high resolution images, led to a decrease in mean absolute error, noting prior vegetation height works had largely ignored textures. Ruiz et al. (2017) argue the integration of spectral, spatial and texture information is vital to earth observation studies particularly in heterogenous urban zones.

## 2.3 Random forests in urban forest mapping

When undertaking forestry research and applying machine learning (ML) methods, the suitable choice of algorithm is vital to the accuracy of results. Different ML algorithms include Support Vector Machines (SVMs), k-Nearest Neighbour (kNN), Neural Networks (NN) and Classification and Regression Tree-based (CART). One of the most popular supervised machine learning algorithms in RS ecological studies are Random Forests (RF) (Breiman, 2001; Mellor et al., 2013). RFs fall within the CART category of ML but differ in that they are an ensemble method of multiple decision trees aiming to reach consensus in classification studies and average in regression. Each tree is formed by taking a bootstrap, using different samples of the training data; whereupon at each node the tree is split by taking a subset of the input variables and the most appropriate variable to split on is determined by GINI index or information gain, a measure of node impurity (Ahmed et al., 2015). RFs have been implemented in many studies due to their many benefits; ability to cope with high-dimensionality data, ability to use heterogenous datasets, ease of model tuning and implementation, resilience to noisy data and no assumption of linearly correlated data (Breiman, 2001; Ahmed et al., 2015; Ruiz et al., 2017).

RFs have been implemented in several forestry contexts in land use classification (Ruiz et al., 2017; Mellor et al., 2013) and urban forest structure studies with great accuracy (Ahmed et al., 2015; Baines et al., 2020; Wilkes et al., 2015). Object based approaches utilizing RFs

have seen application in urban land use mapping (Puissant et al., 2014), and have also been applied to forest structure studies (Wallner et al., 2015; Weinstein et al., 2019) but not in an urban setting.

## 2.4 Google Earth Engine

Google Earth Engine (GEE), launched in 2010, is a web platform capable undertaking large scale computations using analysis ready data and utilising cloud computing processing across several servers (Gorelick et al., 2017; Mutanga & Kumar, 2019). Being a cloud computing platform, this allows for virtualization of super computers for the user, allowing the analysis of "Big Data". Not only does the platform allow for significant computations but also acts as repository for large and longitudinal datasets such as LandSat and Sentinel programmes (Amani et al., 2020).

One of the most popular big geo data processing platforms, the platform allows non expert users to access data through an Application Programming Interface (API) and a web-based Interactive Development Environment (IDE). Further, users do not require expertise in HyperText Markup Language. GEE has ameliorated big data issues in forestry studies; in a study by (Hansen et al., 2013) GEE took 100 hours to process 654,178 LandSat-7 images, or around 707 terabytes to calculate global forest cover and change. The process was estimated to have taken a million hours to complete without the application of GEE.

GEE allows for the export and import of several file formats including .tif and .shp for improved interoperability between platforms. Additionally, several tutorials, scripts and data catalogue are available for users to educate themselves and make research as simple as possible. Another benefit of GEE is the availability of many derivative products. Several popular spectral indices (e.g., NDVI) have in built function. Most of these derivative products are computed on-the-fly upon users' request, due to storage being a larger expense than computation (Amani et al., 2020).

While GEE is a valuable tool, it is not without limitations; on large datasets memory issues can be prevalent causing slowdowns and/or breaks (Gorelick et al., 2017; Amani et al., 2020). Complex machine/deep learning algorithms that require extremely large training datasets cannot be undertaken in GEE, and therefore need to be implemented outside of this platform (Amani et al., 2020). Many studies faced internal errors in parallelizing and executing computations, causing scaling errors and timing out; this likely occurs when analysis outputs

are too large (Tamiminia et al., 2020), additionally restrictions of 250GB capacity may limit some global scale analysis.

GEE has already seen widespread use in urban vegetation mapping (Patela et al., 2015), forest cover (Hansen et al., 2013) and forest structure mapping (Baines et al., 2020). GEE has proven to be a powerful tool for RS with the ability to analyze and classify datasets over differing temporal scales. The continuing improvement of the platform such as the integration within python and continual addition of new datasets, and ease of use provide useful utility for land use and policy makers within the context of urban forests.

## 2.5 Citizen science and i-Tree Eco

Citizen Science (CS) has garnered popularity in recent years to further ecological research and resource management. While urban tree inventories were traditionally undertaken by professional arborists; citizen scientists are now used in many cities. Using citizen scientists in urban forestry has the potential to increase scope of forest studies (Newman et al., 2012), building on tradition of volunteerism in urban forest management (Romolini et al., 2012). In respect to urban forest management, CS projects tend to focus on data collection for tree inventories and monitoring and engagement in strategic planning processes (Roman et al., 2017; Nitoslawski et al., 2019). Although generated data quality is often questioned (Roman et al., 2017), several earth observation projects have used CS to produce high quality data for training and validation datasets (Boyd et al., 2022).

The i-Tree Eco protocol (Raum et al., 2019) is a systematic methodology for urban tree surveying, utilising citizen scientists. i-Tree Eco provides a means to comprehend a city's urban forest measuring its species composition, structure and condition, subsequently calculating and valuing the natural capital benefits that urban forests provide to city residents. The methodology has been used for comparison with modelled data to ascertain model uncertainty (Baines et al., 2020; Alonzo et al., 2016). Baines et al. (2020) noted the discrepancies between CS data and modelled data is likely due to the mode of collection, remotely sensed airborne LiDAR surveys provide a dense set of derived metrics with limited spatial extent, while i-Tree data produce directly measured but sparsely across the entire study area.

Nitoslawski et al. (2019) argues the changing research landscape that recognition that urban forests are "social- ecological" systems, is a turning point for urban forest management.

Whereupon Citizen Scientists should be involved at every level of stewardship including monitoring, maintenance and GI design. Further, a survey of earth observation users by Boyd et al. (2022) revealed urban tree allometry data collection to be one of the few applications in which CS would be trusted. New technologies such as augmented reality could catalyse and empower data collection (Baines et al., 2020; Nitoslawski et al., 2019). Many cities use the citizen science project outputs to improve urban forest management and inform environmental policy (GMCA, 2019). Increased trust and capability of CS projects and Volunteered Geographic Information (VGI) is promising direction for ecological studies.

# 3. Methodology

## 3.1 Study Area

Located within the North west of England (Figure 1), the Greater Manchester (GM) area spans 1280km$^2$ and is home to approximately 2.85 million people (ONS, 2021; Centre for Cities, 2021). Land use is highly heterogeneous, spanning industrial, recreational and residential across 10 local authorities and including the cities of Manchester and Salford. In 2019 $CO^2$ emissions per capita reached 4.3 tonnes, and 28 days with poor air quality in 2020 (Centre for Cities, 2021). ALS missions have been undertaken in the study area from 2010 - 2021, review will be undertaken to assess suitability of ALS datasets to include in the study as estimation accuracy of combining spectral satellite imagery and ALS would be reduced if the fetch interval between the data sources is relatively long.

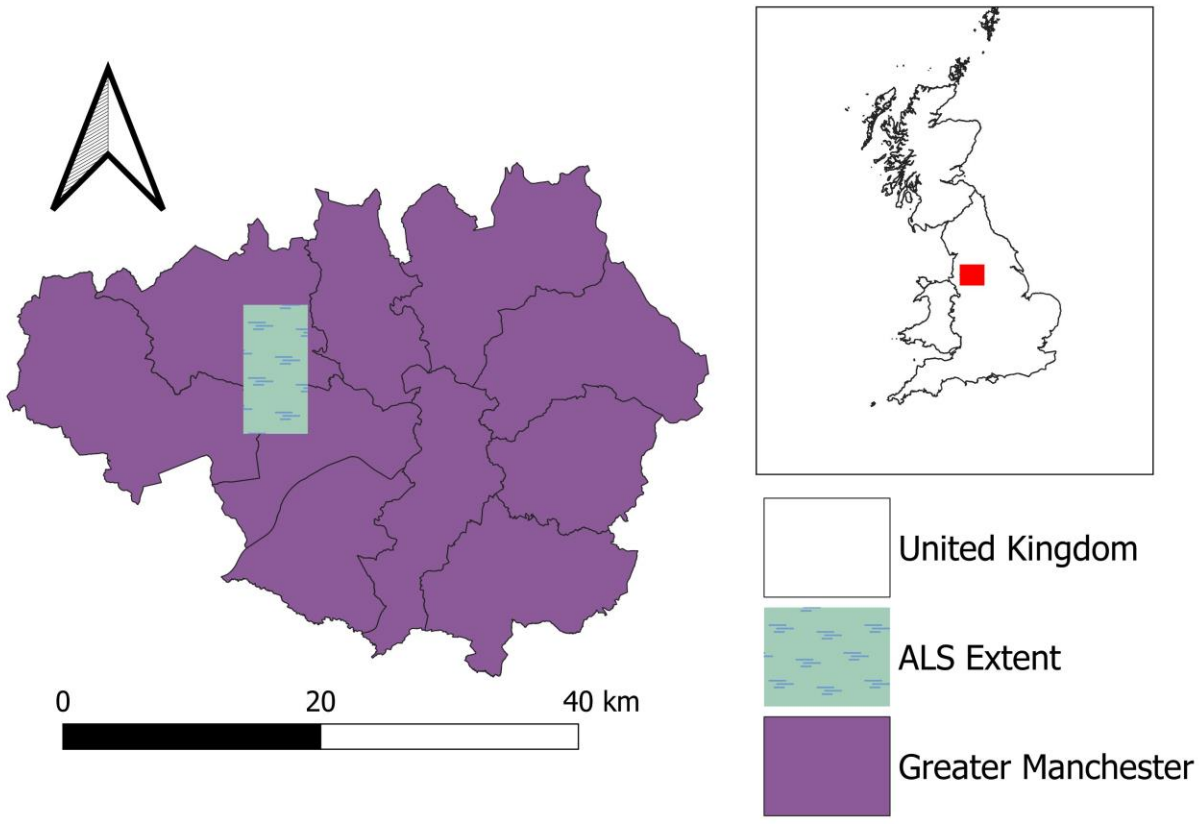



*Figure 1: Greater Manchester Study area highlighting aerial LiDAR survey coverage*

Urban forest research is of considerable interest in the UK, with the research arm of the Forestry Commission; Forest Research previously and currently undertaking several studies to ascertain urban forest structure in towns and cities across the UK (Doick et al., 2016).

Potentially the largest physical survey of trees in the UK; the Greater Manchester i-Tree Eco project undertaken in 2018 estimated a total of 11,321,386 trees with 15.7% of GM under canopy, broadly similar to the UK average urban tree canopy cover of 16%. Tree density equates to 88 trees per hectare, significantly higher than London (53 trees/ha) and the UK town average (58/ha) (CITY OF TREES, n.d.). The report indicates that distributions are unequal, areas of dense population equate to low canopy cover. The Greater Manchester Combined Authority (GMCA) and charity Manchester City of Trees have set ambitious targets as part of a five year environment plan for GM to plant three million trees by 2035 and a further one to two million by 2050 (GMCA, 2019). It is worth noting unlike canopy cover targets, planting targets do not account for tree losses, further GMCA debates the usefulness of CC targets stating "While we are keen to understand the extent of Greater Manchester's canopy area and, ultimately, increase this area, new planting takes several years to form a mature canopy and means that change can only be monitored over longer timescales. Also, there is no target, whether it is numeric or aerial, that represents 'enough" (CITY OF TREES, n.d.). Finally, the Northern Forest Project aims to plant over 50 million trees in the North of England encompassing Greater Manchester, with the aim of creating £2.5 billion in ecosystem services (Woodland Trust, 2022). This study intends to present a framework to allow stakeholders to economically monitor the success of such strategies.

## 3.2 Preprocessing LiDAR data

The UK Environment Agency (EA) undertakes Airborne LiDAR Surveys freely available in 5km tiles as part of the National LiDAR Programme with surveys starting in November 2016. Tiles were downloaded in .laz format from the UK Government's data portal (Open Government Licence v3.0). Although freely available, pulse density is generally low, ranging from 1 – 2 points per metre squared, limiting the size of tree that can be detected.

Tiles covering the study area were collected from missions undertaken during the summer season of 2018, to coincide with i-Tree Eco ground surveys and increase likelihood of capturing canopy cover in summer season. Individual tiles were mosaicked to form a continuous layer using QGIS software (QGIS Development Team, 2022). Canopy height model creation and tree crown segmentation was undertaken using the LidR package (Roussel & Auty, 2021) in R. Tree crowns were aggregated to points, producing summary measures: CC, CH and N. A total of 97,835 trees were identified.

To examine differences in model performance two resolutions were included for analysis. Firstly, a coarse resolution of 100m to reduce computational power and therefore computing constraints. Second, a medium resolution of 20m as to utilise the resolution of Sentinel 2 bands 5,6,7,11 and 12. It is intended to improve spatial detail in comparison to 100m outputs, however if little significant difference is produced between datasets, then this method may be more applicable in wider settings, and applied to lower spatial resolution imagery.

Two regular grids, 20m and 100m were intersected with the vector crown layer. CC was computed as the sum of the 2D projected crown area, where overlapping crowns were dissolved and then divided by grid square area to find the proportional canopy cover per grid pixel. CH was calculated from the $95^{th}$ percentile of heights per grid square to ascertain dominant CH. Tree number was calculated as the crown centroids per grid square, using centroids ensured crowns overlapping grid squares were not counted multiple times.
To reduce the likelihood of spatial autocorrelation within the response variables a random sample of the 20m point layers was taken, leaving 10,000 datapoints. This was deemed unnecessary for the 100m resolution dataset given the coarseness, thereby leaving 5,000 datapoints.

## 3.3 Preprocessing Sentinel 2 data

Google Earth Engine (GEE) hosts several datasets including the Sentinel series of satellites. A cloud computing platform operated by Google providing options for data visualisation and analysis (Gorelick et al., 2017). In this study GEE has been used as part of a multistage process, to compile datasets then subsequently download for further analysis, this method has been used effectively in machine learning studies (Hird et al., 2017), however machine learning forest inventory mapping has also been undertaken on the cloud (Duan et al., 2019a). Offering the advantage of no requirement to download, store and manage analysis ready data (ARD) and given the ever expanding data leverage gap and issues pertaining to big EO data, the GEE platform follows a trend of open cloud computing (Hird et al., 2017).

Sentinel 2 imagery was accessed and preprocessed to a Level-2A surface reflectance, correcting for atmospheric distortions. common format, projection (OSGB 1936) and spatial resolution. A Sentinel 2 Level-2A cloud free imagery was computed from images captured between 01/06/2018 to 30/06/2018. A cloud mask was created masking values with a QA band of anything outside 0 (analogous to clear conditions) using GEE tutorials. Landsat-8 imagery was considered as part of this study however Sentinel-2 imagery was chosen due to the higher spatial resolution, 30m to 10m respectively. Additionally, both spectral resolutions are comparable and both being available on GEE.

Although gap filling algorithms exist to ensure the continuous coverage of EO imagery, GEE allows for the creation of image composites very easily. Image stacks can be created from image collections in the GEE catalogue to form one composite image. After the application of masks Sentinel 2 images were filtered from the period of June 2018, stacked together then the median value taken from the pixels. As the median value is resilient to outliers this average was chosen. This created an image with 13 spectral bands.

Additionally, a water mask was applied to remove bodies of water where no trees reside using the Modified Normalised Difference Water Index (MNDWI) (Xu, 2007; Du et al., 2016) given these areas are known to be free of urban trees, incorrectly assigning structure metrics to these areas could impact the results. MNDWI was preferentially chosen over Normalised Difference Water Index (NDWI) as this index incorporates green and Short Wave Infrared (SWIR) bands, this diminishes built up area features that can be correlated with open water in other indices (Xu, 2007). MNDWI equates to:

*Equation 1*

$$MNDWI = \frac{(Green - SWIR)}{(Green + SWIR)}$$

Although Xu (2007) suggests a default threshold of zero, a subsequent threshold value of 0.15 was chosen after a qualitative visual assessment using Google Earth imagery.

## 3.4 Reflectance bands and ancillary variables

Additional ancillary variables to compliment surface reflectance were also derived. Texture metrics including variance and mean were computed for each image in 3x3 image windows, to illustrate the changing reflectance intensity across a different scale. Further, given the

importance of temporal information specified within the literature review, another Sentinel 2 image stack was composited, spanning a 2 year period. Variance in pixel values was calculated for this period for each spectral band, due to phenological changes throughout this period in deciduous vegetation it is expected to show a larger variance value for these pixels, whereas non vegetated surface such as the urban fabric would exhibit a lower spectral variance.

NDVI was a key variable included in the study to discriminate between urban and vegetated areas, in addition to account for phenological change. NDVI was calculated using the following formula:

*Equation 2*

$$NDVI = \frac{(Near\ Infrared - Red)}{(Near\ Infrared + Red)}$$

Further, to compliment surface reflectance texture metrics, NDVI texture metrics of variance and mean were calculated in 3x3 and 5x5 windows. Temporal variance of NDVI was calculated across a two year span to further represent change over time.

GEE also hosts several other complimentary datasets other than satellite imagery. Ancillary information that could impact vegetation growth was also collated including elevation, slope, and aspect taken from Shuttle Radar Topography Mission (SRTM) and precipitation and climatic data from the WORLDCLIM1 project (Hijmans et al., 2005). All predictor variables were reprojected to British National Grid (OSGB 1936) and then resampled to 20m and 100m. The complete list of 47 predictor variables is in Table 2 for both spatial resolutions.

*Table 2:Predictor variables for Random Forest models*

| *Input variable* | *Source* | *Acquisition date* |
|---|---|---|
| *Surface Reflectance (SR) Bands 1-9,11 and 12*<br>*SR Band mean (3x3)*<br>*SR Band Variance (3x3)*<br>*NDVI*<br>*NDVI mean (3x3, 5x5)*<br>*NDVI variance (3x3,5x5)* | Sentinel-2 MSI: MultiSpectral Instrument, Level-2A | 1st June 2018 – 30th June 2018, 15th April 2021 – 15th June 2021 |

| | | |
|---|---|---|
| *NDVI temporal variance* | Sentinel-2 MSI: MultiSpectral Instrument, Level-2A | 1st January 2018 – 31st December 2019 |
| *Elevation*<br>*Slope*<br>*Aspect* | Shuttle Radar Topography Mission, NASA SRTM Digital Elevation 30m | 11th February -12th February 2002 |
| *Mean Annual Precipitation*<br>*Mean Average Temperature*<br>*Maximum Average Temperature*<br>*Minimum Average Temperature* | WorldClim Climatology V1 | 1st January 1960 – 1st January 1991 |

## 3.5 Selecting predictor variables

As aforementioned, RFs are popular in ecological research due to their resilience to outliers, not requiring normally distributed data and being able to cope with correlated datasets. It was recognized that in some machine learning models the issue of collinearity; can confound and lead to inaccurate results. Highly correlated variables when combined can impact linear models, however as a RF regressor machine learning model is applied here, model results are unaffected by colinear variables, the principal purpose of this study. As such all-predictor variables as part of the feature engineering process were kept in the model. Nevertheless, RF feature importance will still be impacted by non-independent collinearity (Dormann et al., 2013; Breiman, 2001).

## 3.6 Random Forest machine learning

Ensemble CART-based models RFs use the voting of several independent trees to form a consensus, the majority vote in classification tasks and average in regression tasks. This should lead to the benefit of not overfitting the data, a problem associated with other model types and reducing the impact of noisy data. To build each independent tree in a RF generation of a random subset of the dataset is taken, thereupon a subset of input features are examined to ascertain where which results in the largest decrease in error or most information gained, to determine which splits occur at each tree node. Given the stochastic nature of RFs results can vary between forests as dependent on which variables are chosen to split upon. Bootstrap sampling of inputs can be applied to improve the generalisation of RFs.

RF was applied within R (R Core Team, 2018) using the *randomForest* package (Liaw & Wiener, 2002), utilizing the regression mode as to predict the urban forest continuous variable metrics. Data was split 70:30 into training and validation sets respectively. For the 20m resolution model 10,000 points were randomly selected as a compromise between performance and processing times while still greatly exceeding the training sample requirement specified by (Mather, 2011) of 30 points per input variable. RF models were trained using LiDAR data and then combined and applied to the input data to predict values for cover, height and tree number across GM.

Hyperparameter tuning of the number of trees within each forest ,*ntree* and *mtry*, the number of random variables to test when splitting at each node was undertaken (Mellor et al., 2013) to reduce Root Mean Square Error (RMSE) and improve accuracy. Hyperparameter tuning was undertaken using the *caret* package (Kuhn et al., n.d.), For *mtry* a grid search of 1 to 30 variables was undertaken using a k = 10 fold cross validation, intended to limit and reduce overfitting of the dataset, while being a compromise on computing time. For each tuning process a fixed number of 100 trees was limited to each forest. This process produced resulted in producing the *mtry* value with the lowest RMSE within the 1:30 range for each 20m/100m model. To tune *ntree* the optimal *mtry* for each model was taken to be used in another grid search using 100,250, 300, 350, 400, 450, 500, 550, 600, 800,1000 trees and executed to ascertain the lowest RMSE value. These tree values were chosen to be an acceptable compromise between computation time and theoretical decrease in error. Hyperparameter tuning results in the optimal number of trees for each RF, shown in Table 2 which were then adopted into the final models.

*Table 3: Hyperparameters chosen after model tuning.*

| *20m* | *mtry* | *ntree* |
|---|---|---|
| *CC* | 14 | 600 |
| *CH* | 2 | 800 |
| *N* | 7 | 1000 |
| *100m* | | |
| *CC* | 14 | 600 |
| *CH* | 9 | 800 |
| *N* | 20 | 1000 |

Finally, RF were then run using optimal parameters using the randomForest package using a 70:30 training/test split. Once RFs had produced outputs, each individual pixel was mosaicked together to form a continuous output of urban forest structure metrics. Change detection was undertaken by calculating the pixel difference in R using the overlay tool.

## 3.7 Comparison against i-Tree Eco

Ground inventory data was collected from the Manchester i-Tree Eco project ('i-Tree Eco | City of Trees', 2018), one of the largest i-Tree projects outside of the United States, utilizing a team of 57 surveyors and visiting nearly 2000 field plots, each 0.04 ha in size. Analogues for ALS derived structure metrics from each plot concerning CC ("Percentage Tree Cover",), CH ($95^{th}$ percentile tree height) and N (number of trees within plot) were extracted for

comparison. Points containing i-Tree Eco data were overlayed onto maps of RF outputs, for N, values predicted by RF were extracted for the intersecting i-Tree plots. The Methodological steps are shown in Figure 2.


*Figure 2: Methodological design*

# 4. Results

## 4.1 Model performance

The greatest variance explained within testing sets are the 100m N ($R^2$ = 0.7501) (RMSE = 8.02). The lowest variance explained is the 20m CH ($R^2$ = 0.32) (RMSE = 6.14). Model performance $R^2$ and RMSE metrics are presented in Table 3. Performance metrics indicate lower error in CC and CH for 100m RF model, while N has lower error for 20m model.

*Table 4: RF model error metrics for 20m and 100m scale.*

| | *r²* | *RMSE* |
|---|---|---|
| *20m* | | |
| *CC* | 0.6366 | 13.75 |
| *CH* | 0.3285 | 6.14057 |
| *N* | 0.4547 | 0.952729 |
| *100m* | | |
| *CC* | 0.7501 | 8.020622 |
| *CH* | 0.3399 | 5.637549 |
| *N* | 0.73 | 9.788645 |

Residual values indicate somewhat weak relationships with variable predicted value, as shown in Figure 3 and 4 For the 20m  CC plot, there appears to be a n error relationship where error grows until around 40% , whereupon negative error decreases, while positive error remains large.

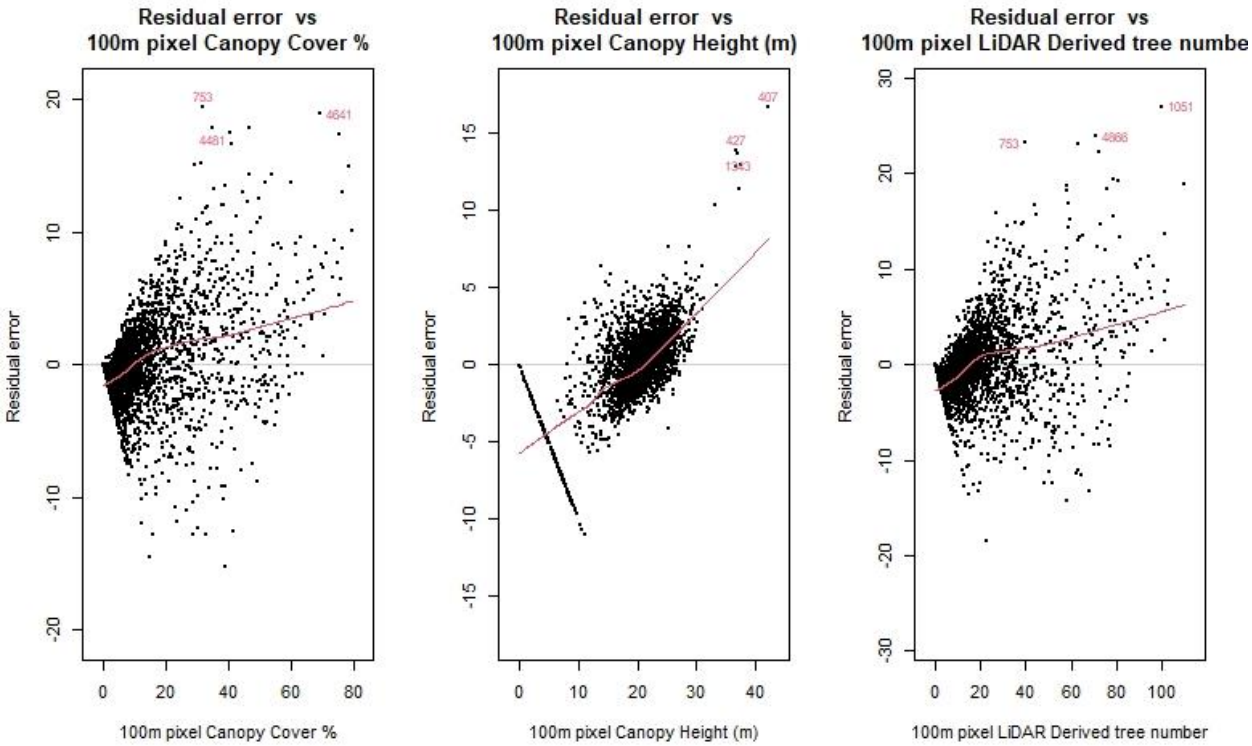



*Figure 3: Residual model (100m) errors plotted against forest structure metrics.*

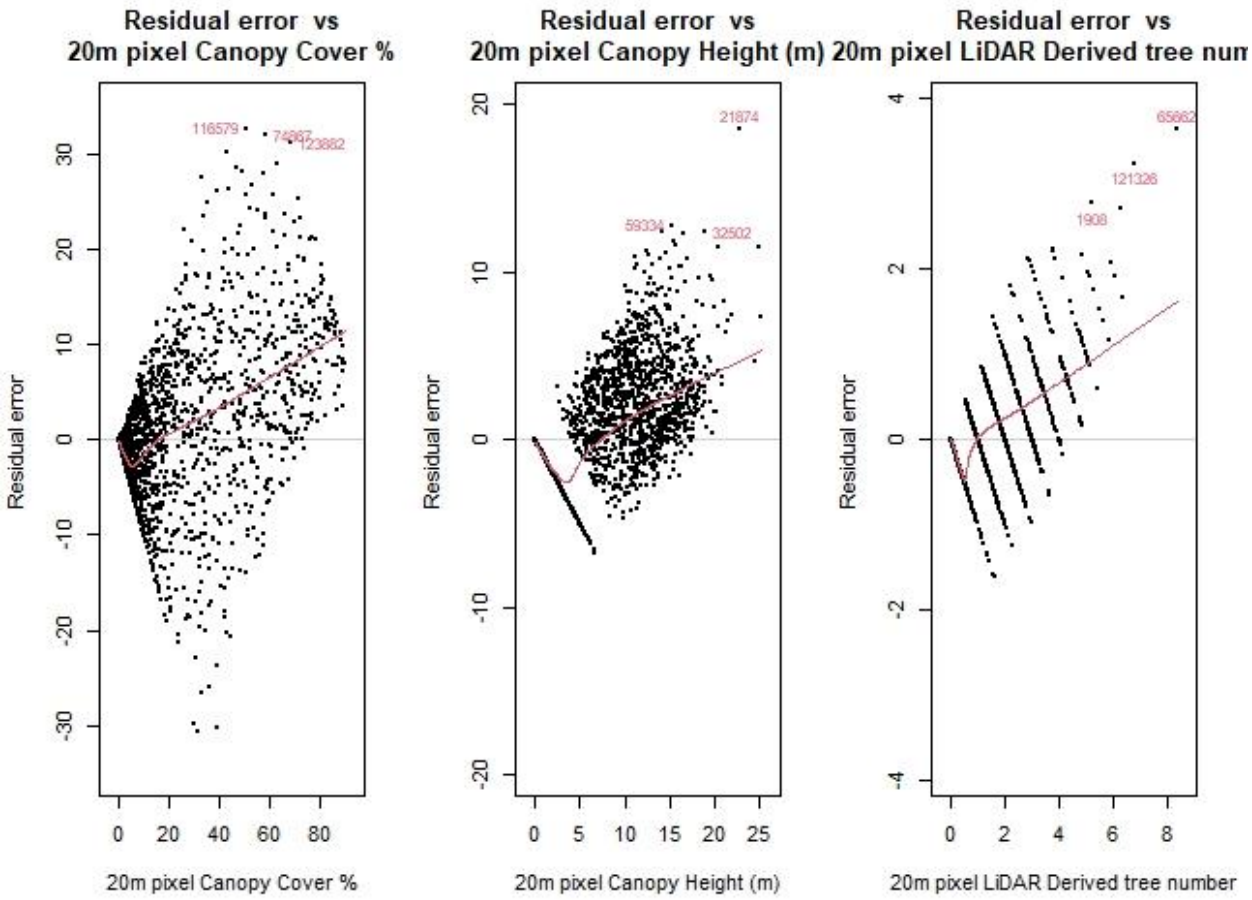



*Figure 4:Residual model (20m) errors plotted against forest structure metrics.*

Figure 5 and 6 indicate the how the surface reflectance variables are most prominent in explanation of variance within each model. Both figures show how for every model other than 20m CC and N, Sentinel 2 Band 5, collecting data in the Vegetation Red Edge range was the most important predictor variable for 100m CC,CH and N, 20m CH and N, and third most important for 20m CC and N. This is particularly notable in the N models where removal of Band 5 from the 100m N model would increase Mean Square Error by 117%. Also prominent in the models is Band 3 green, in the visible spectrum, with %MSE values of over 200% in 20m CC, and 91% for 100m CC. The NDVI predictor variable also have a significant contribution to variation within all 20m and 100m scale models, appearing all six models tops 10 important variables; particularly when predicting CC in the 20m model accounting for over 76% %MSE, and 40% %MSE for N (100m). Band 4 representing the visible red spectrum, also features prominently, particularly in the 20m models, where for N it is the most important band and for 20m CC, accounts for 141% %MSE.

Of the remaining variance, texture metric Band 5 mean, appears in all six models top 10 important variables, and is second most important (6.9% %MSE) for 100m CH. However, surface reflectance metrics appear more important than their texture counterparts, where in all the models top 10 importance values, Band 3, 4 and 5 are all more important relative to their texture metric.

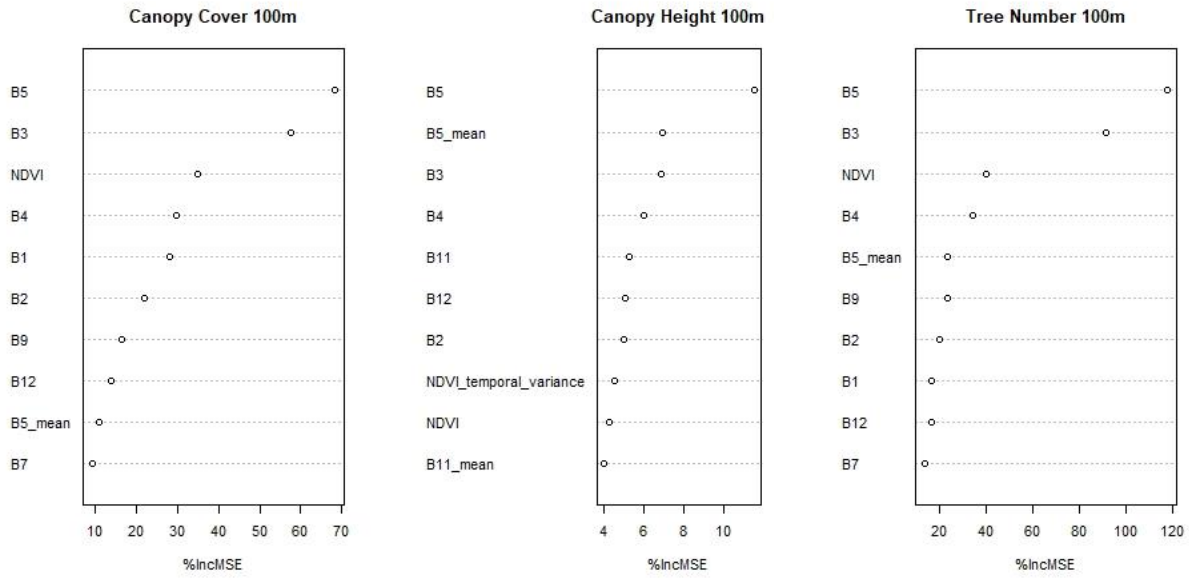



*Figure 5:Variable importance plots for 100m models.*

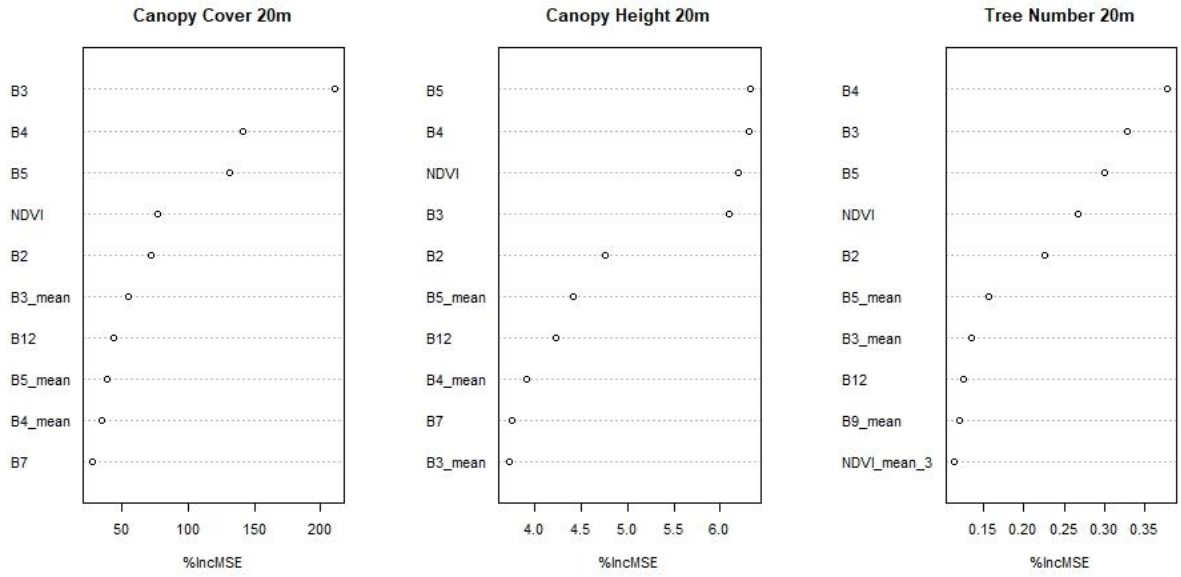



*Figure 6:Variable importance plots for 20m models.*

## 4.2 Urban Forest Structure

97,835 trees were derived from EA LiDAR surveys within the study region, after LiDAR processing in R. Maps of CC, CH and N are presented in Figure 7. In total 2701238 trees have been located within GM in 2018, equivalent to 0.84 trees per pixel using 20m trained RF. This value is significantly lower than the 2018 i-Tree Eco model estimates of 11,321,386 (CITY OF TREES, n.d.) and 2010 Bluesky values (4,794,857) taken from the National tree map extracted from aerial photography using photogrammetric techniques (BLUESKY, 2017). Mean CC was found to be 14.99 %, while CH has a mean height of 18.8m. Table 4 shows statistics of the derived structure metrics.

*Table 5: RF urban forest structure metric results*

| | *Min* | *Max* | *Mean* | *Std Dev* | *Sum* |
|---|---|---|---|---|---|
| *100m 2018 CC (%)* | 0.01510914 | 79.417 | 14.987 | 11.799 | n/a |

| | | | | | |
|---|---|---|---|---|---|
| *2021 CC (%)* | 0.39653352 | 69.08 | 14.234 | 8.961 | n/a |
| *2018 N* | 0 | 110 | 20.475 | 14.368 | 2638883 |
| *2021 N* | 1 | 87 | 19.641 | 10.35 | 2529879 |
| *2018 CH (m)* | 0.12007867 | 42.13145 | 18.845 | 4.184 | n/a |
| *2021 CH (m)* | 2.37756968 | 30.70992 | 19.104 | 4.1492 | n/a |
| *20m* | | | | | |
| *2018 CC (%)* | -5.14E-14 | 91.166 | 14.835 | 17.455 | n/a |
| *2021 CC (%)* | -5.03E-14 | 83.042 | 13.296 | 12.863 | n/a |
| *2018 N* | 0 | 8 | 0.8462 | 0.9444 | 2701238 |
| *2021 N* | 0 | 5 | 0.8187 | 0.7297 | 2611026 |
| *2018 CH (m)* | 4.56E-02 | 25.23 | 5.6771 | 3.7811 | n/a |
| *2021 CH (m)* | -5.17E-15 | 17.094 | 5.8082 | 3.0768 | n/a |

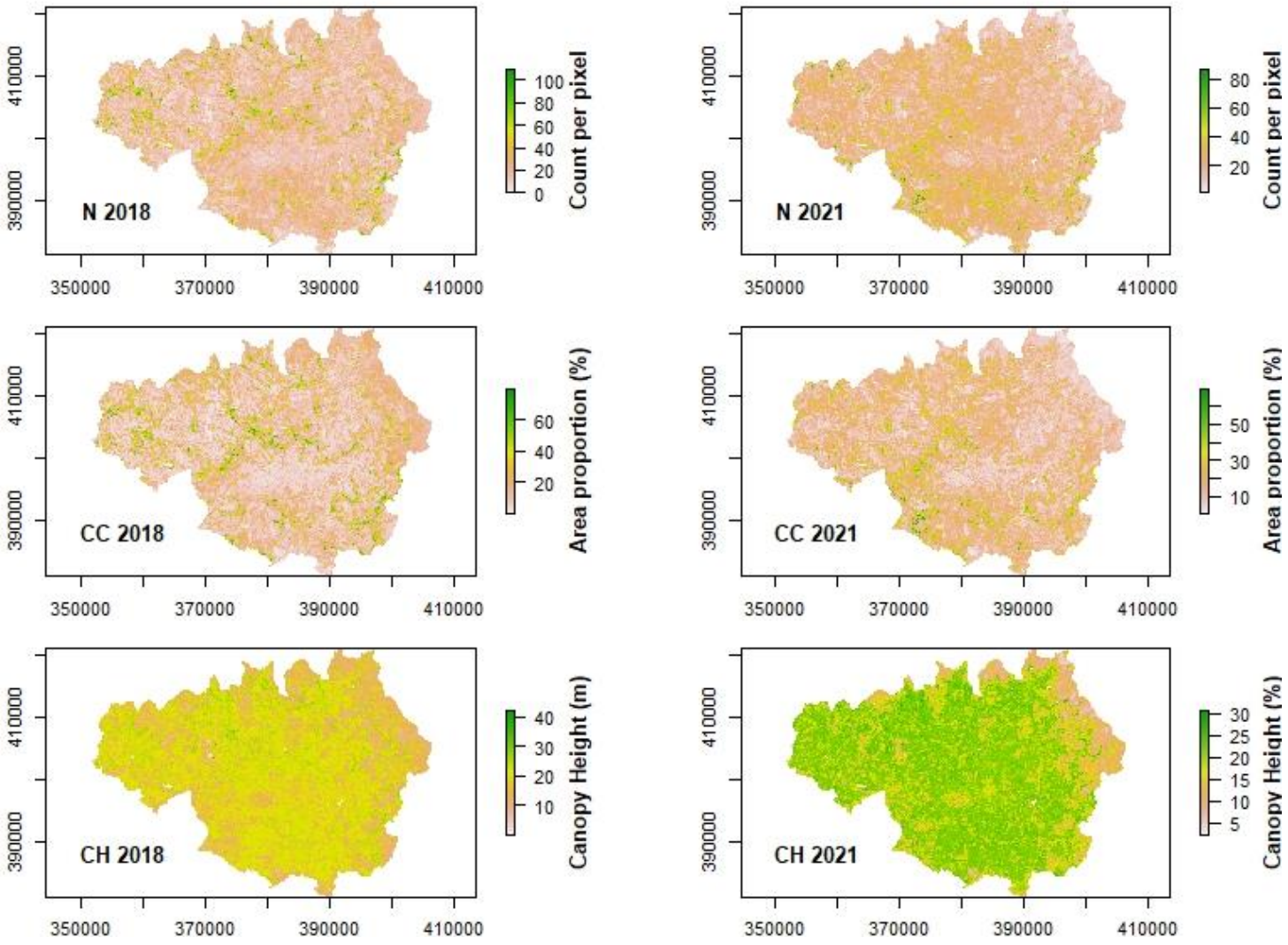



*Figure 7: Forest structure metric maps estimated through 100m random forest.*

Total summed CC derived from the 100m model was 19316.6ha, accounting for 15.1% total CC, slightly below the national average of 16% (Doick et al., n.d.), and below 2010 Greater Manchester Tree Audit (16%) (CITY OF TREES, n.d.). Figure 7 illustrates the spatial spread of forest structure across GM, while Figures 8 and 9 illustrate the frequency distributions of both 20m and 10m RF models and ALS derived estimates.

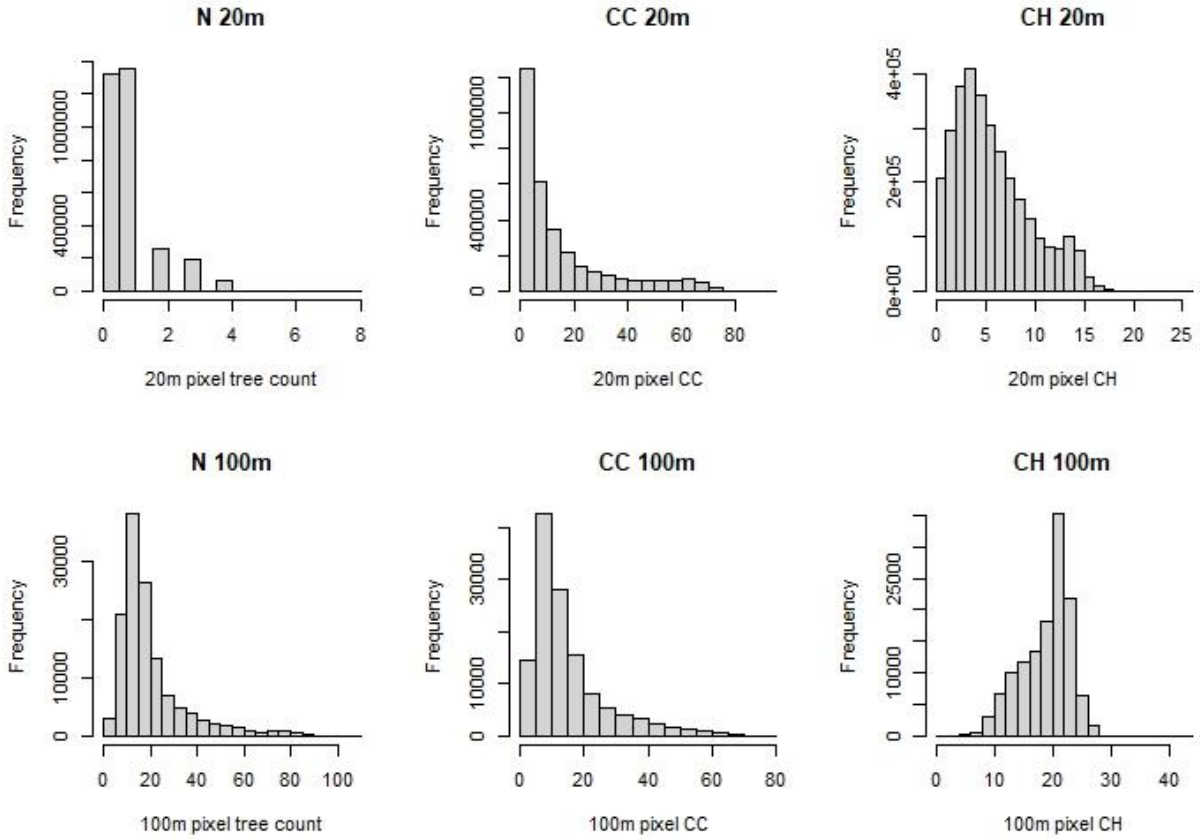



*Figure 8: Frequency distributions of RF modelled 2018 forest structure metrics.*

RF models appear to overestimate values of CC, CH and N compared to ALS derived estimates, this is thought to occur due to ALS training data being skewed to the left, however a feature of RFs is a regression to the mean, reducing skewness, therefore leading to an overestimation and underestimation in continuous values (Wilkes et al., 2015). Correction functions are available in the randomForest package, however correction functions are often indiscriminate and can lead to incorrect scaling and increases in overall error values (Baines et al., 2020). Figures 10 and 11 illustrate the relationship between forest structure variables at 20m and 100m scale, with distance from the centre of GM.

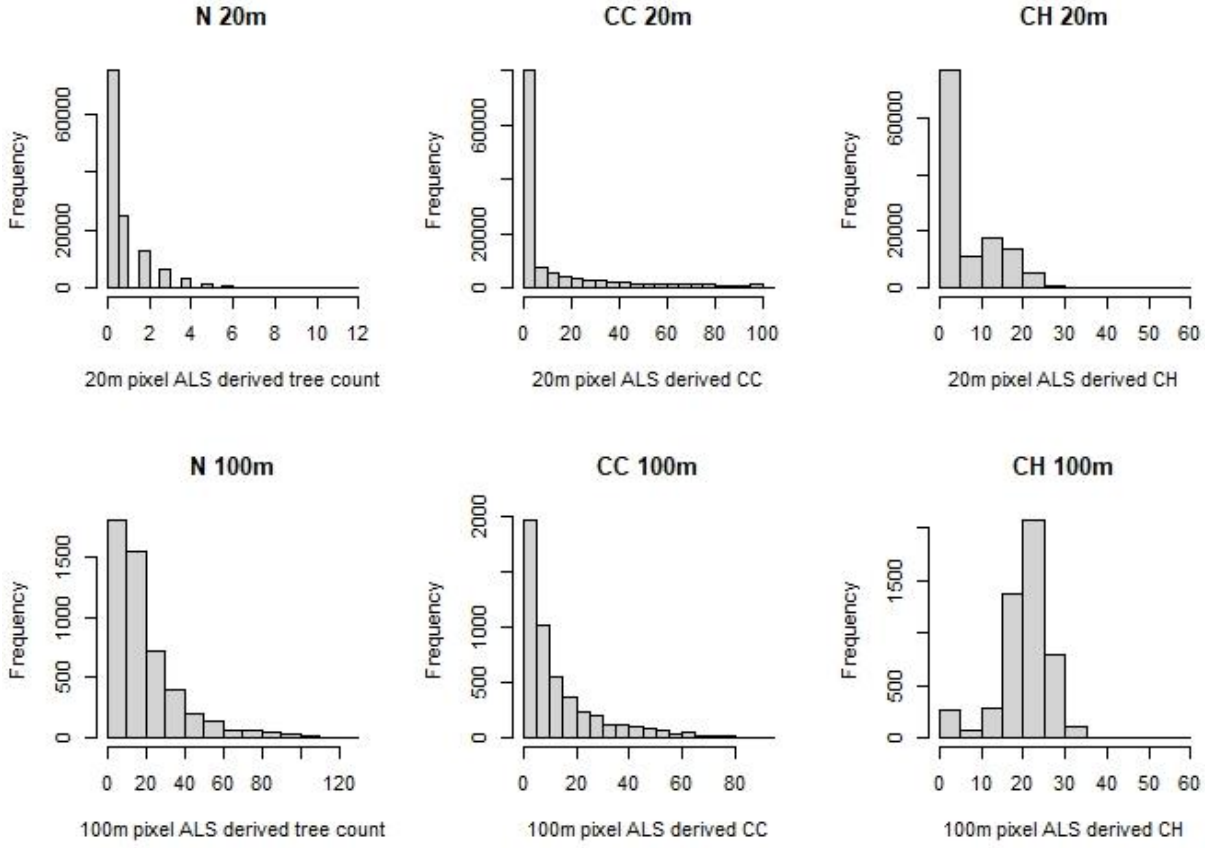
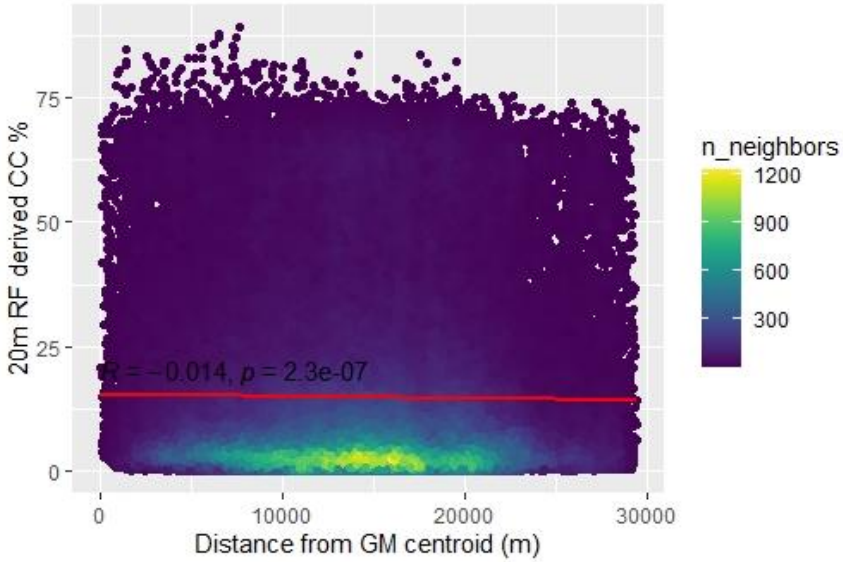
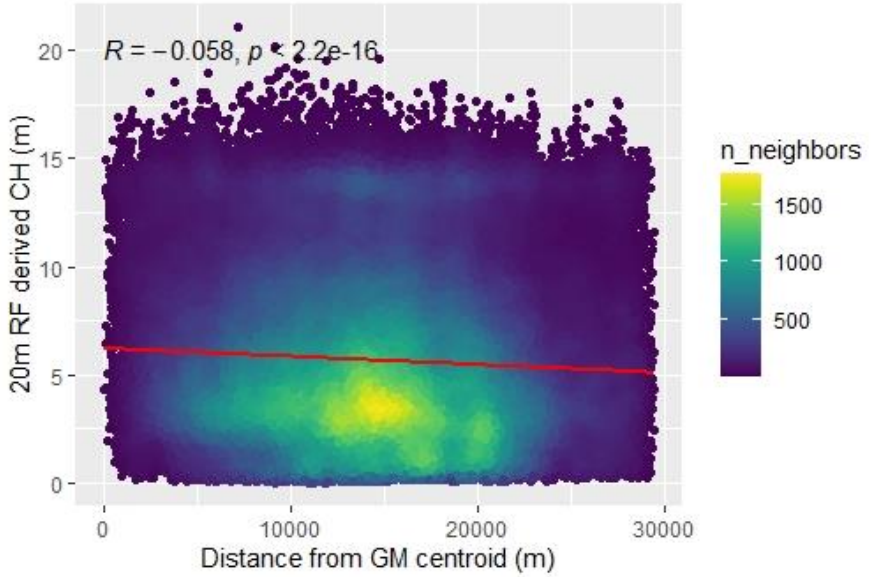



*Figure 9: ALS derived forest structure metrics*



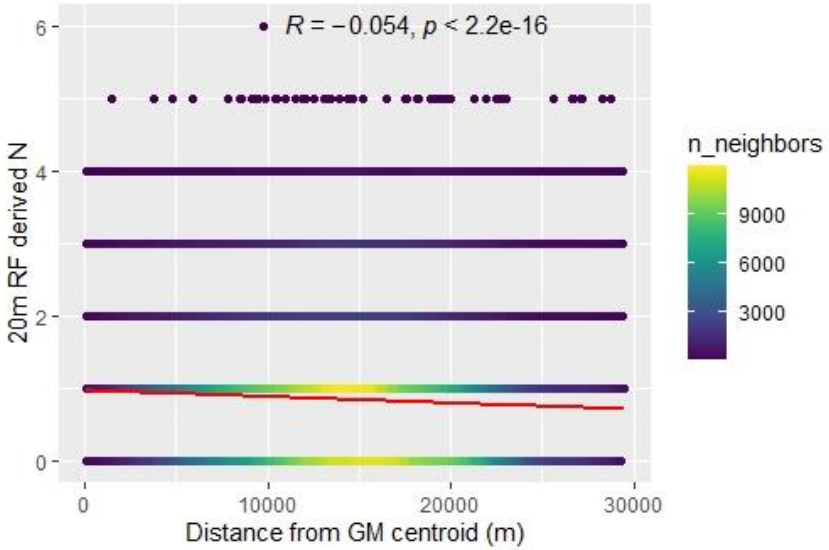




*Figure 10:Scatterplots of tree number (N), Canopy cover % (CC) and Canopy Height (CH) derived from 20 RF plotted against distance from centre of GM.*

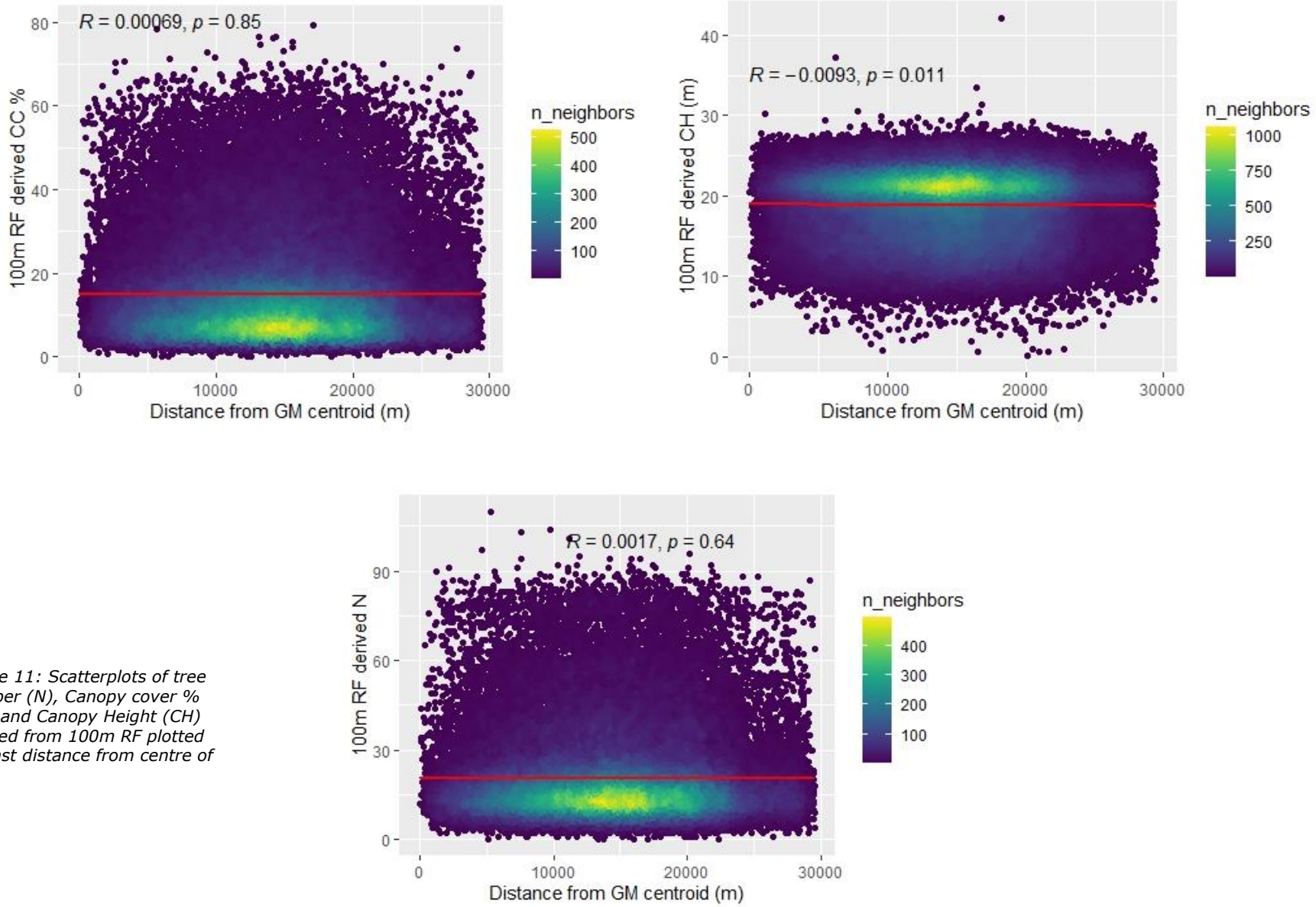



*Figure 11: Scatterplots of tree number (N), Canopy cover % (CC) and Canopy Height (CH) derived from 100m RF plotted against distance from centre of GM.*

The Number of trees per plot, N and canopy height (CH) did not increase with distance from the centre of GM, for the 20m RF model, both maximum and minimum values are distributed across GM, with a weak negative correlation with distance of R = -0.054 for both metrics. Similar trends can be seen across GM for metrics at both 100m and 20m scales. However, for the 100m scale model results for N and CC exhibit positive relationships with distance R = 0.0017 and R = 0.0069 respectively.

Between variables, correlation is equally strong between CH and CC, and CH and N (r = 0.92), while very slightly weaker for CC and N (r = 0.91); and all correlations were all highly significant ($p<0.001$), Figure 12 illustrates the relationships between variables derived from the 20m RF.

Figure 13 plots the borough level estimates for the 20m RF derived forests structure metrics, illustrating both 2018 and 2021 years. The plot indicates the Wigan district hosts the most trees, an estimated 392874, while the borough with the least trees is Tameside district with 207397 trees. Figure 13 also displays average CC across boroughs, the borough with the highest mean CC is Stockport District with 16.04% CC, while the lowest district is 12.77% CC corresponding to Rochdale district. Finally Figure 13 shows average CH across boroughs, where the district with the largest average CH is the Manchester District with an average height of 6.25m; the district with the lowest average CH is Oldham district with a value of 5.17m.

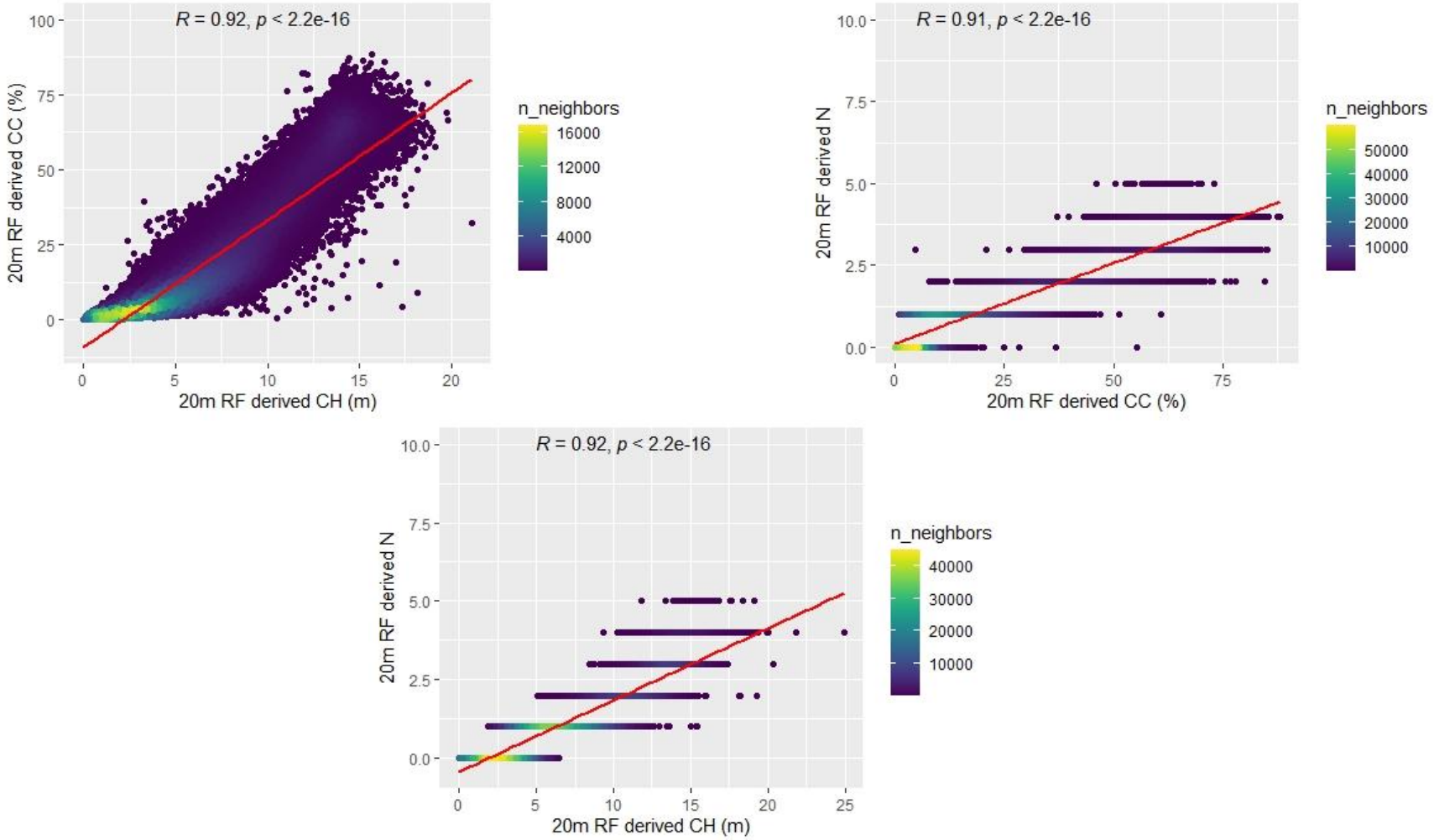



*Figure 12: 20 RF derived forest structure metrics plotted against each other.*

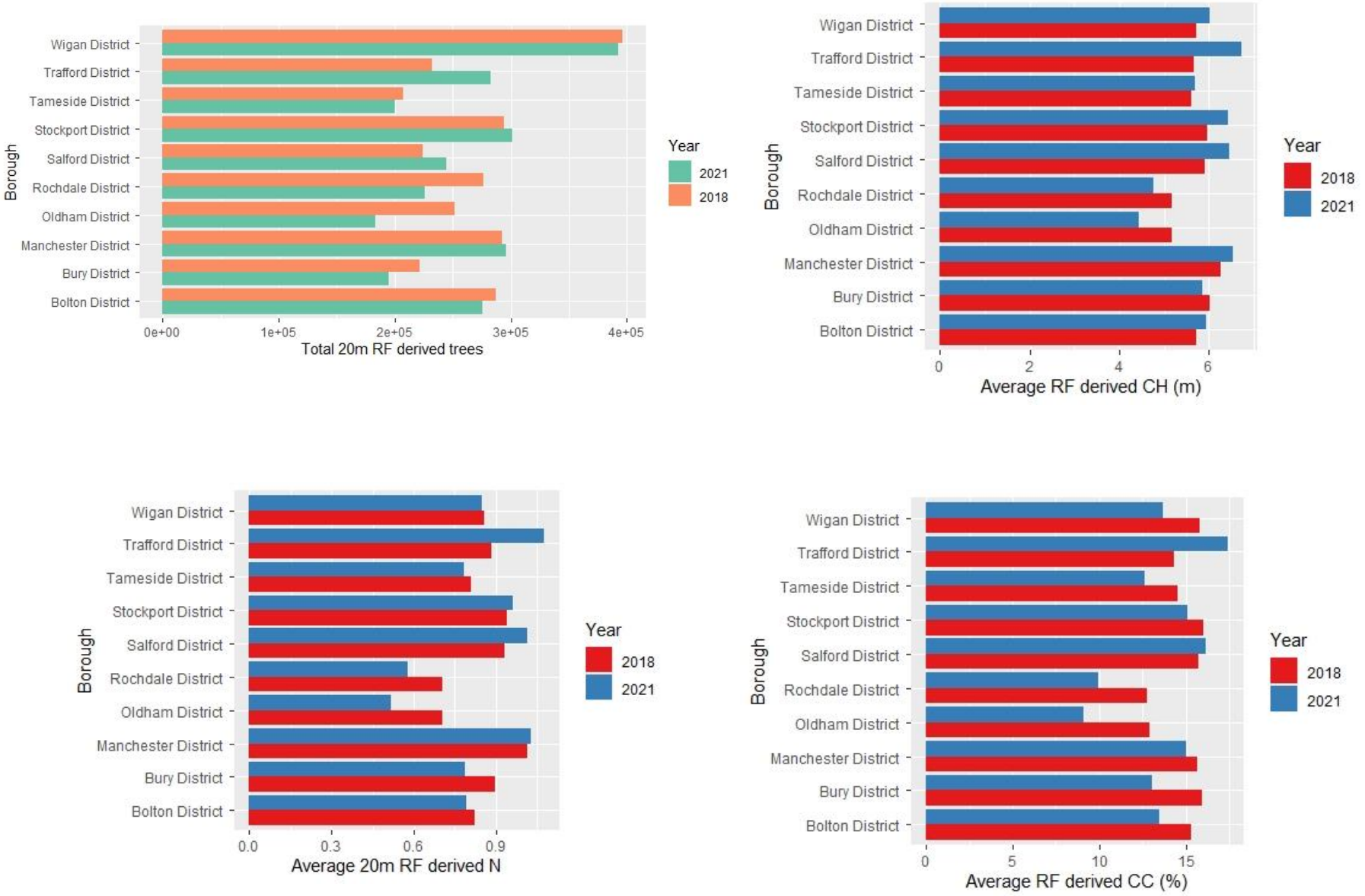



*Figure 13: Borough level estimates of average 20m RF derived CC CH, N and total number of trees.*

## 4.3 Resolution change

Tree number estimates derived by the 100m RF model were smaller in comparison to the 20m RF, estimating around 2.6 million trees at 100m and 2.7 million at 20m in 2018, and around 2.5 million compared to around 2.6 million for 2021 data (Table 3). However, estimates for CC, were reversed when the 100m RF estimated slightly higher values. Mean CC for 2018 applying the 20m RF was 14.99%, compared to 14.84% in the 100m RF in the same year, and this trend was shown again for 2021 derived data. This pattern is also shown in CH, where mean CH in 2018 for 20m RF was 5.67m compared to 18.85m for 100m RF, this trend continued into 2021 data.

The model overall variance explained was greater for the 100m model compared to the 20m, for all metrics: CC ($R^2$ = 0.75), CH ($R^2$ = 0.34) and N ($R^2$ = 0.73). However, this does not completely correlate to RMSE values. RMSE values for 20m N are smaller (RMSE = 0.95m) than the equivalent 100m (RMSE = 9.7m). frequency distributions are illustrated in Figure 9. Examining 2018 data, CC was more positively skewed for 100m than 20m (skewness = 1.75 and 1.66 respectively), this trend was also shown in N with skewness = 1.31 and 2.03 for 20m and 100m respectively. However, in comparison of CH between scales, CH for 100m had a skewness value of -0.66, while CH for the 20m RF had a positive skewness of 0.81.

## 4.4 2018 against 2021

Figure 14 illustrates the forest structure metric change from 2018 to 2021 over the entirety of GM. Total tree number decreased by nearly 109004 trees in the 100m map, and 90212 in the 20m map. This accounts for a mean tree number change per pixel of -0.83 for 100m and -0.02 for 20m pixels, maximum tree changes per pixel were 85 for 100m and 5 for 20m. Similar trends are shown in the mean CC change, both decreasing in the 20m and 100m by -1.5% and -0.75 respectively. Further, maximum values for CH decreased in both 20m and 100m RF models from 2018 to 2021, the largest decrease being in the 100m RF, of 10.m. However, mean CH however increased by 0.26m in the 20m map and by 0.13m in the 100m map.

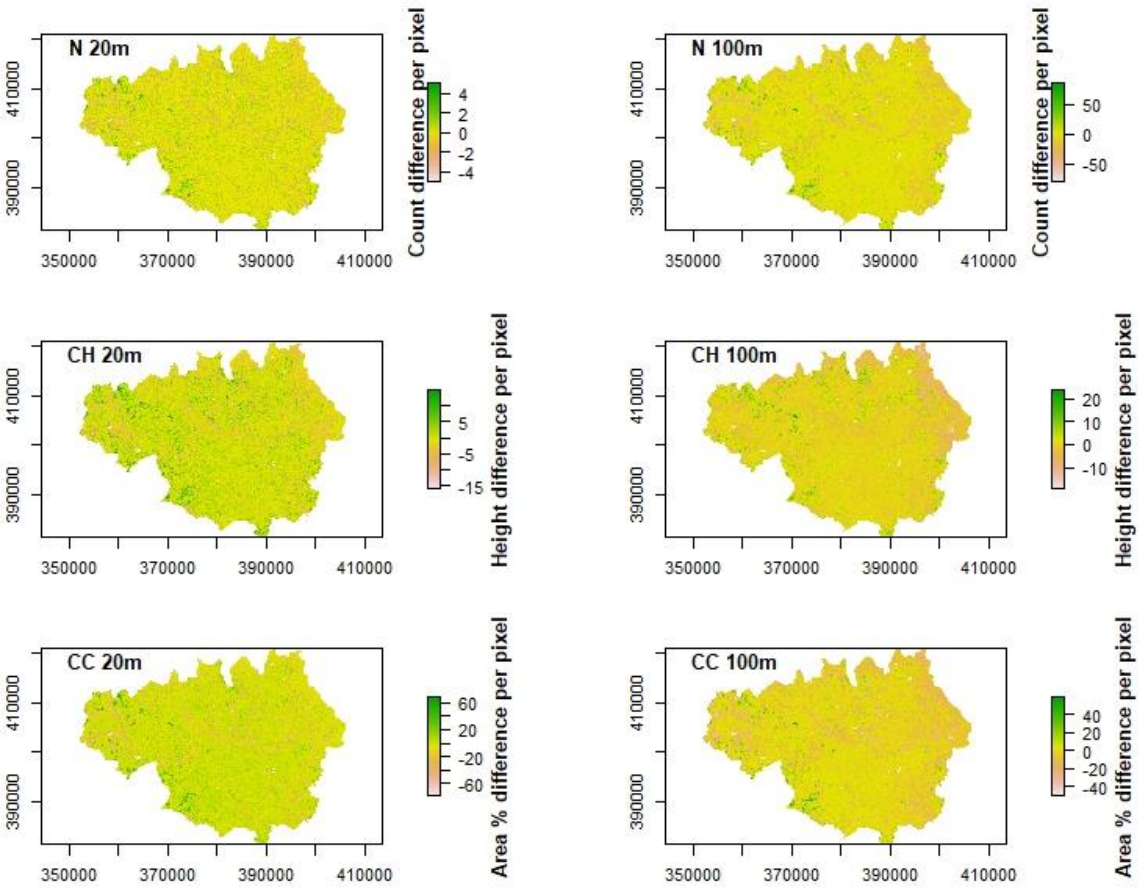



*Figure 14: forest structure difference plots 2018 to 2021*

Figure 14 illustrates the changes in forest structure between 2018 and 2021. The charts indicate towards the centre of GM, change of forest structure metrics are lower than areas towards the edge GM, for both positive and negative alterations.

Figure 15 illustrates the change of forest structure between 2018 and 2021 are not uniform between boroughs throughout GM. Additionally, changes between structure metrics are not equal between each other and across differing scales. Total tree number difference appears to indicate a general agreement between 20m and 100m scales and illustrates across GM all boroughs gained or lost trees during this period, where six out of ten boroughs experienced tree losses. Manchester district indicates the smallest change with an increase of 3605 and 4146 trees for 20m and 100m scales respectively. However, the largest change in tree numbers occurs in Oldham District, where a large decrease of -68031 and -64040 trees took place for 20m and 100m RFs. Some boroughs also experienced large increases in tree

numbers, Trafford district indicated a sizable increase of 50,317 and 45,759 trees estimated by 20 and 100m models.

Figure 15 also illustrates how trees per pixel changed throughout the time period. Similarly, the Trafford district produces the largest increase in trees per 100m pixel (+4.33), while Oldham displays the most significant decrease in mean trees per 100m pixel (-4.5). Stockport District exhibits the smallest absolute mean change, with a mean change of only 0.32 trees per 100m pixel.

Mean CC change per pixel, is also described below, when applying the 20m RF, only two boroughs experienced increases in mean CC, Salford and Trafford districts, results from the 100m RF also indicated CC gain in Manchester borough, showing disagreement between models. The largest increase in mean CC per pixel is Trafford borough with an increase of 3.8% for the 100m and 3.12% for the 20m, Salford has a smaller increase of CC, 1.36% for the 100m and only 0.48% for 20m. The largest decreases in mean CC over the time period are shown to have occurred in Oldham District, decreasing by -4.4% and -3.84% for the 100m and 20m RF respectively. The smallest loss of CC occurred in Stockport for the 100m RF (-0.26%) and Manchester for the 20m RF (-0.66%).

Finally mean CH change was extracted per district, demonstrating that mean CH has been dynamic between 2018 and 2021. Seven out of GMs ten districts exhibited loss of average CH according to both 20m and 100m RFs, however discrepancy exists where the 20m RF identifies the Bury district as having a decrease in mean CH, while this decrease is identified in Tameside for the 100m RF. The 100m and RF describes the largest differences in CH as occurring in again Trafford, a mean increase of 1.2m; compared to the largest decrease of -1.25m occurring in Oldham district. The 20m RF agrees with the larger scale model, where Trafford hosts the largest increase compared to other boroughs, a mean increase of 1.1m, and with the largest decrease again occurring in Oldham, -0.73m. Tameside appears to have the least changed average CH, changing by only -0.10m in the 100m RF.

Figure 15 illustrates the spatial heterogeneity and non-uniformity of urban forests structure across GM, changing on between regions and within scales. Trends appear to agree with large structural changes in Oldham and Trafford districts, expressing large changes of both negative and positive magnitude.

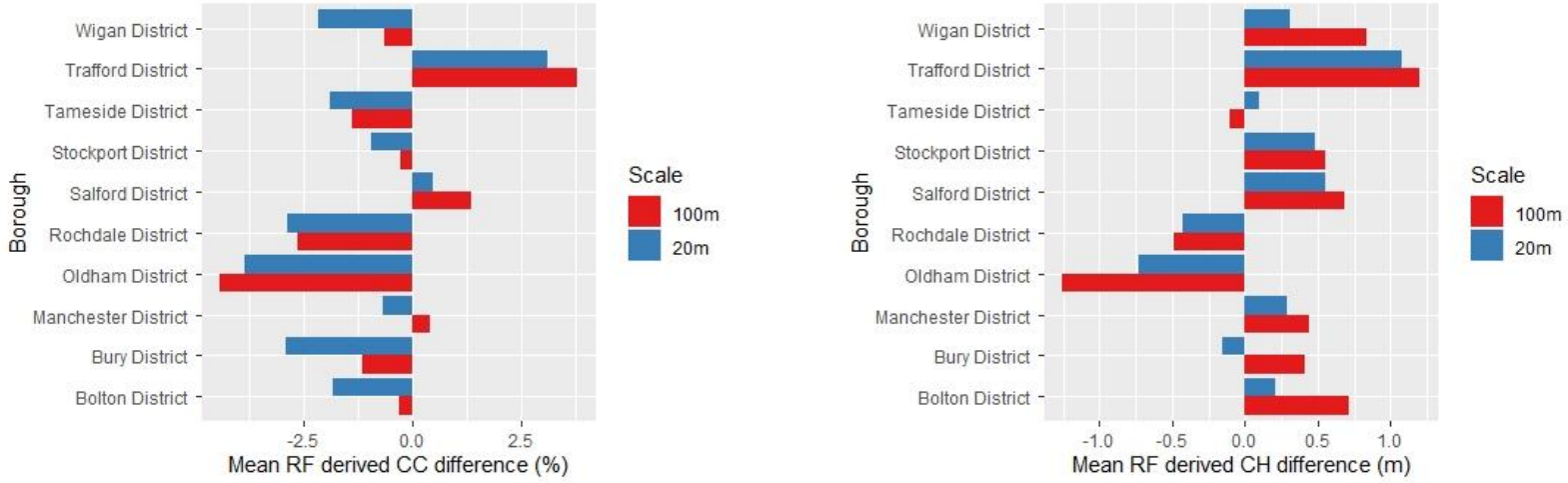
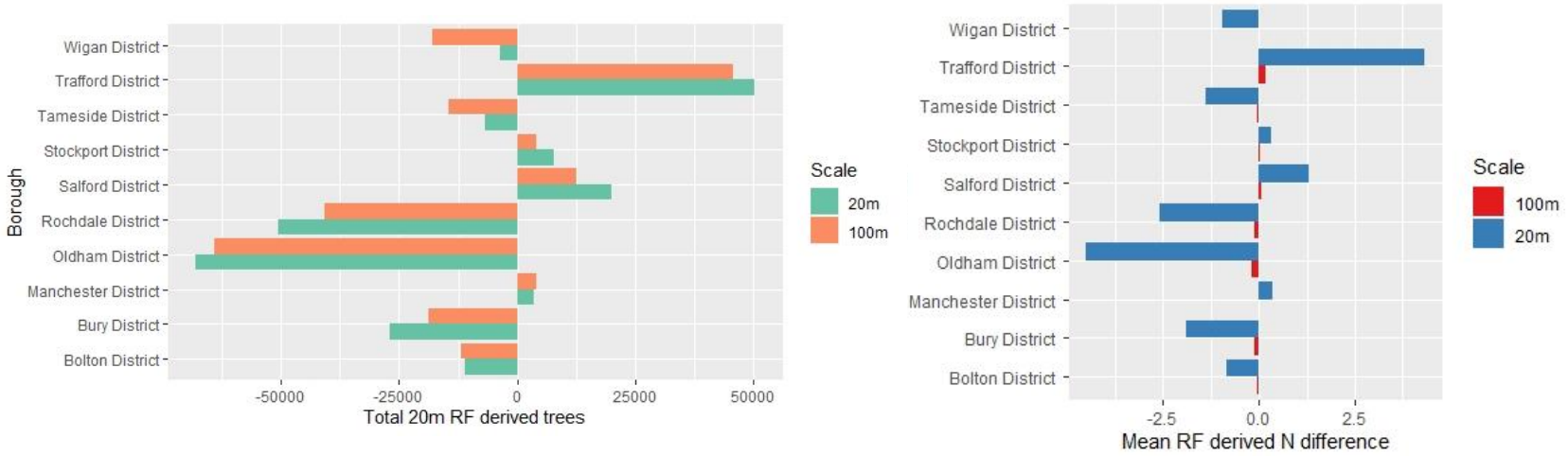



*Figure 15: RF derived forest structure metrics differences between 2018 and 2021 for both 20m and 100m scales.*

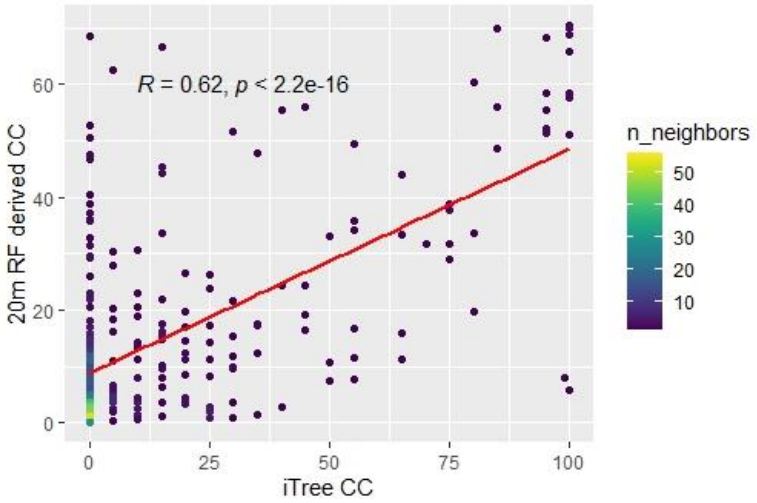


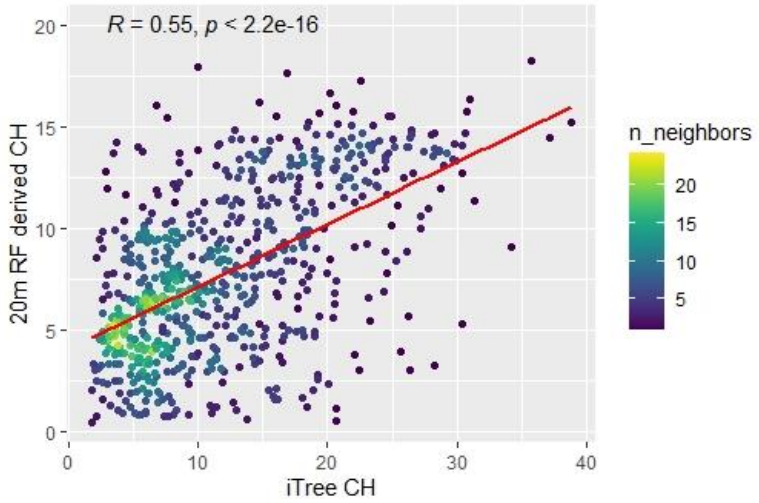
## 4.5 Comparison with i-Tree Eco

Comparison with i-Tree Eco plot level data, (number of plots = 1876) indicates somewhat disagreement (Figure 16). RMSE is 21%, 14.9m and 7.6 trees for CC, CH and N respectively. Of the 1876 plots 666 plots had tree counts of one or more, while 1210 plots has no tree identifications associated. Whereas, 4,197 plots from the ALS derived area were estimated to be treeless, or 60% of the 20m derived plots. Comparing derived CC and i-Tree CC data illustrates a medium strength relationship, R =0.62; it appears however CC is tends to be overestimated by RF as indicated in Figure 16. CC data collection using the i-Tree Eco framework uses a visual interpretation for assessment, and producing values in a range (i.e. 10-15%), this differs from ALS derived CC detection. CH from RF and i-Tree has a weaker relationship, R = 0.55, again appearing to over estimate CH compared to ground data. This discrepancy is likely caused by ambiguity of the germination point, analogous to the centroid of the crown envelope (Baines et al., 2020). This discrepancy is particularly obvious where, i-Tree plots are found to be treeless but RF outputs derived CC as over 60%, this occurs where large trees with extended crowns overlap neighbouring plots and leading to difficulty in understanding satellite data (Baines et al., 2020). Several, i-Tree Eco plots, display CH and N, however lack corresponding CC data, suggesting tree bases are present but their canopy lay outside of the plot boundary, leading to a difference in data collection techniques between ALS and i-Tree fieldwork.

A weaker relationship is exhibited between i-Tree Eco plots and RF for N, R = 0.46; 20m RF appears to significantly underestimate N particularly in highly tree dense plots, tree number within 20m pixels. Maximum N for i-Tree Eco data is 81 trees, whereas the maximum RF estimated was 5 per pixel in the intersection and only a maximum of 8 per 20m pixel across GM, a hugely significant difference , suggesting the predictor variables have difficulty in accurately resolving individual trees.

Total modelled i-Tree Eco GM forest structure metrics, published in Manchester Tree and Woodland Strategy,(CITY OF TREES, n.d.), produces a figure of 11,321,386 trees with 15.7 per cent of Greater Manchester beneath tree canopy. Total tree number is significantly greater than 20m RF estimates of 2,701,238 trees, a difference of 8,620,148 trees in 2018. Average CC shows a closer relation, 14.8% and 15.1%, a difference of 0.9% and 0.6% for 20m and 100m RF models respectively.



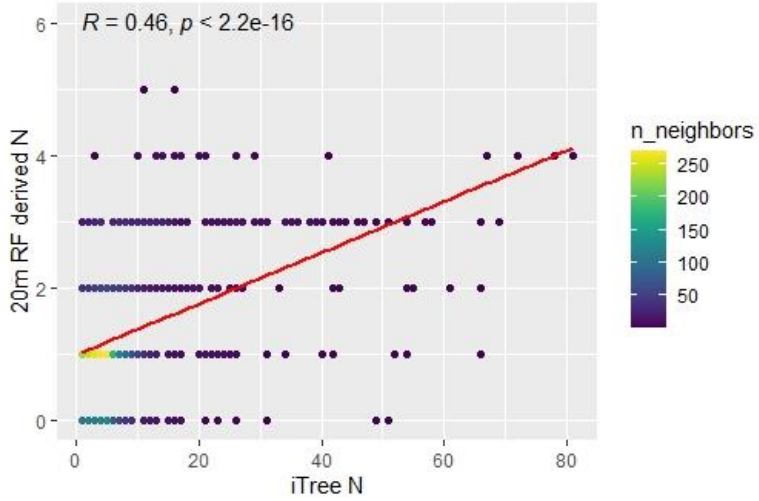




*Figure 16: Scatterplots of 20m RF forest metrics against i-Tree Eco plot level data.*

i-Tree Eco plot frequency distributions are presented in Figure 17. In comparison with figure 8, frequency distributions of 100m and 20m RF forest structure metric estimates, all metrics present positive skewness 1.90, 2.58 and 0.76 for CC, N and CH respectively. Plot level data for i-Tree Eco data reveals stronger positive skewness for all derived metrics from the 100m RF, and all metrics in 20m RF results other than CH, with a skewness of 0.81. However in comparison to ALS derived metrics, skewness is larger in both CC (2.01) and CH (1), indicating in general i-Tree Eco data is more right skewed than RF estimates but less than ALS estimates.

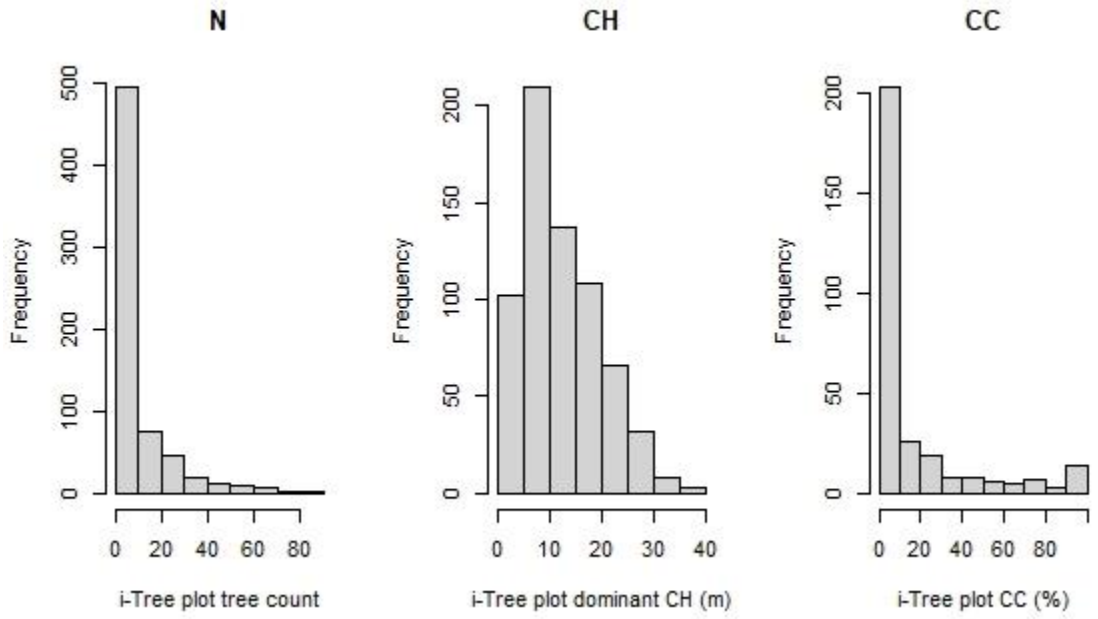



*Figure 17:i-Tree Eco forest structure frequency distributions.*

# 5. Discussion

## 5.1 Model performance

The moderately high variance explained within both the 20m and 100m testing datasets for CC, where 64% and 75% of variance was explained, lends credibility to the derived CC output. Multispectral imagery can detect variation across land surface, thereby delineating between vegetation/canopy and anthropogenic structures through varying spectral signatures. Variance explained is likely attributed to the capability of physical spectral contrast. CC RMSE values of 13.75% and 8.02% for 100m and 20m RF respectively is a significant improvement over CC estimates for Oslo undertaken by (Hanssen et al., 2021) of 32.6% when using only LiDAR segmentation methods. Whereas Duan et al., (2019b) applied Sentinel 2 imagery, ground data and RF for urban forest mapping across China with errors with around 7% error, notably smaller than this studies' estimates. Applying LiDAR, Sentinel 2 imagery in an RF, to map urban forest structure in Southampton and Greater London; Baines et al. (2020) produced RMSE values ranging from 11-17% correlating with results from this study. However, compared to spatially homogenous studies utilizing LiDAR and multispectral LandSat data, where Ahmed et al. (2015) produced RMSE values of 0.0714, across rural Canada, this studies results appear relatively large. Given the spatial heterogeneity in the above studies, RMSE values produced in this study appear to agree with (Parmehr et al., 2016; Li et al., 2019), that fusion of LiDAR with satellite imagery leads to reduced error in CC detection for urban studies.

Variances for both N and CH, were lower across both 20m and 100m RF models, compared to CC. For N 45% (20m) and 73% (100m) variance was explained in the test sets. For CH much lower 33% (20m) and 34 % (100m) variance was explained, a much lower value than Wilkes et al., (2015) attained 54%. RMSE values of CH in this study are comparable to Wilkes who produced an RMSE value of 5.7m in 2015, and then RMSE ranging from 4.9–6.2 m for London forest mapping in 2020 (Baines et al., 2020). However again compared to a homogenous forest environment RMSE values are relatively larger (Ahmed et al., 2015).

For trees per pixel, N, various canopy sizes for input values ranging up to three orders of magnitudes up to a maximum value of 3835 $m^2$, suggesting issues with segmentation, but also indicates the spread of canopy sizes per tree. In contrast CH value ranges will likely have reduced detection by RS due to little horizontal space being encroached as CH increases. Therefore, an increase in CC may not lead to an increase in tree number, as a large number

of trees may account for large or small canopy cover and vice versa. As a result, it is possible the spectral contrast may be reduced, across canopy, segmentation may be improved by leveraging supplementary datasets such as age, or species to stratify forest stands (Ahmed et al., 2015).

All output datasets were positively skewed with the exception 100 CH, which illustrated a negative skewness. Strong positive skews for CC and N indicating mean values are greater than median. Smaller skewness value for CH which may explain the lower RMSE error.

Importance variables in the RF models presented in Figures 5 and 6, help illustrate the dominance of surface reflectance variables within the models, in particular red, green and vegetation red edge bands. Similar to Baines et al., (2020), where red edge and SWIR were most important across structural metrics. Red edge importance is due to the large response from vegetation in this wavelength (670-760 nanometres), important for retrieving chlorophyll content information and leaf area index (Delegido et al., 2011).

NDVI is also an important variable shown in Figures 5 and 6, this spectral indice has been shown to be an important variable for classification and regression RS ecology studies (Wang et al., 2010; Duan et al., 2019b; Parmehr et al., 2016). NDVI has the ability to improve accuracy by identifying LiDAR tree points in the segmentation process, resolving from other urban infrastructure.

Texture metrics have been applied in other studies of forests structure (Baines et al., 2020; Lang et al., 2019), and in agreement several texture mean metrics appeared in the top 10 variable importance ranking, but not as significant as surface reflectance in most cases. Decreased importance of texture metrics is likely due to the increased spatial heterogeneity or larger presence of anthropogenic features than natural forest systems, for example in CH mapping in rural Australia by (Wilkes et al., 2015) found texture to be highly important, compared to Baines et al., (2020) , where CH was mapped in London, and texture metrics played a less important role.

Wilkes found temporal metrics of NDVI to be significant, this is also in shown in our 100m RF CH model, where NDVI temporal variance is placed above NDVI for explaining CH distribution. The expression of phenological changes over time, are recorded by this metric, however different species of trees and vegetation, providing a varying phenological response in larger

pixel sizes may confound results, in mixed stands, potentially explain why this metric does not serve a higher importance in the other models.

Notably, the topographic and climatic variables are exempt from the top importance ranking, although this agrees with Baines (2020), other studies into forest classification have found these variables to be important, where lower land is associated with decreased precipitation but also clearing for agricultural land (Mellor et al., 2013).

## 5.2 Urban Forest Structure

Wall to wall maps of urban forest structure metrics at 20m scales provide an improved spatially explicit approximates of 3D urban forest structure and distribution, aiding in the future analysis of higher spatial resolution data in the urban fabric. This dataset could be particularly pertinent given many urban forests are overlooked in global estimates of ecosystem services and may even rival magnitude and variety of benefits compared to natural forest stands (Wilkes et al., 2018).

Visual comparison with 2018 UK Centre for Ecology & Hydrology (UKCEH) land cover maps shows general agreement with canopy cover and tree number maps, in delineation of woodland areas, illustrated by Figure 18 and 19. RF estimates can help identify many parks across GM, including Chorlton Water park (Figure 19), Drinkwater and Philips Park in Prestwich, and Botany bay wood to the west of Eccles. Although RF estimates are useful in identifying macro level trends across GM, discrepancies compared to other datasets such as Bluesky's NTM and i-Tree eco modelled data suggest methodological issues. This likely attributed to the derived LiDAR dataset being unrepresentative of the entire GM region. Mean ALS derived tree number per 20m pixel was 0.78 trees, compared to an average 9.23 per 0.04ha i-Tree Eco plot, taken from thousands of plots spread across GM. Conversely, in line with the i-Tree methodology, where plots are located in areas of very high tree density overestimates can occur as higher than average density is then extrapolated to the entire study area.

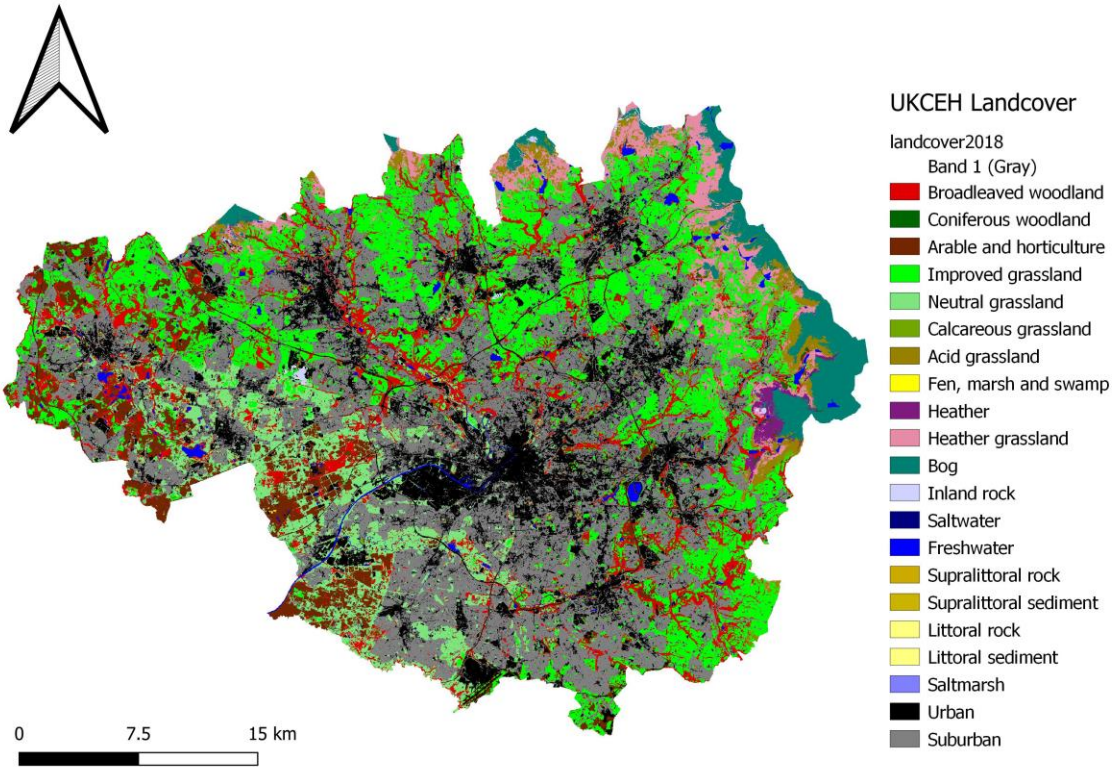



*Figure 18: UKCEH Landcover map for GM.*

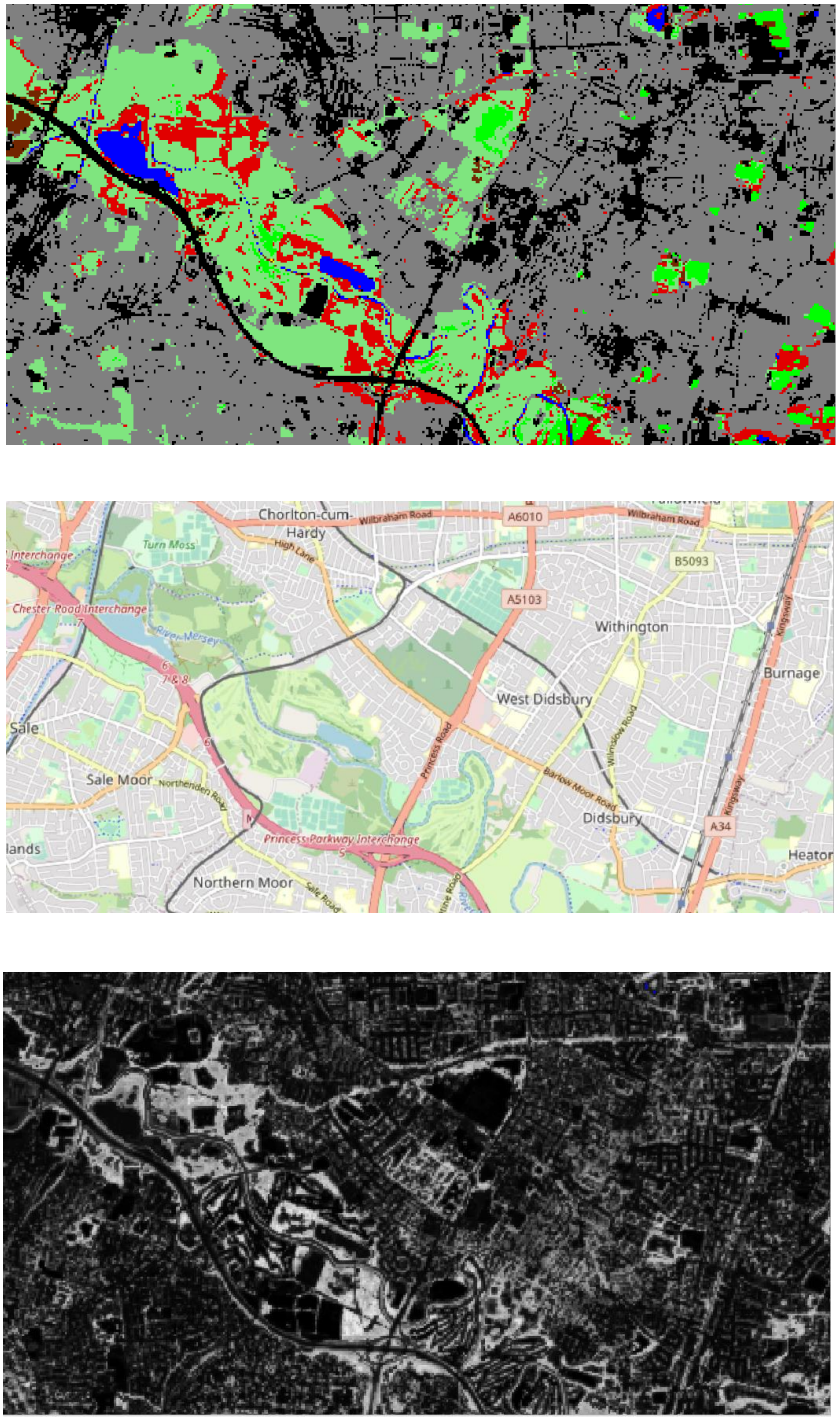



*Figure 19: Visual inspection of Chorlton park, from 2018 UKCEH land cover map (top) where red represents woodland, OSM (middle) and 2018 CC estimation, where white areas represent high canopy cover.*

The role of urban trees will continue to play a valuable role in ensuring metropolitan areas, home to around 10 billion trees (Endreny, 2018) remain habitable and resilient to shocks to society in the wake of climate and population changes. Inventory tools to first collect information on quantities of natural capital, and vital for future management and maintenance, and subsequently a method of assessing effectiveness of urban forest management policy decisions. Correctly informing decisionmakers is crucial to preserve these natural assets and maximise derived ecosystem services.

This study presents a framework, based upon the work undertaken by Baines et al. (2020), utilizing RF methods to estimate wall to wall 3D forest structure metrics in a complex heterogenous urban setting, across spatial and temporal scales. Applying open access and freely available datasets imbues UKRI FAIR principles and creates opportunity for application of this method in other urban regions. Given the modular nature of the framework, separate datasets could inform each stage to suit researchers; ALS could be replaced with TLS, spaceborne LiDAR or i-Tree Eco plot structure data. Predictor variables taken from Sentinel 2 and supplementary sources, could be acquired from elsewhere such as the LandSat series. Additionally, interchanging appropriate software based upon researcher competency, such as including ArcGIS products or Python may lower perceived barriers for further studies. Modularity in this approach should ensure flexibility for researchers.

As noted by Baines et al., (2020) the approach should not be considered as replacement for fieldwork inventories but to augment current practices, this is particularly pertinent given the disparity between i-Tree ground data and predicted metrics. Medium RMSE values of 21 %, 14.9m and 7.6 trees indicate the immaturity of the method, and requirement of further refinement. Further, fieldwork inventory protocols capture supplementary information useful for ecosystem assessments including species composition, which may exceed the capacity of current RS technology or may be obscured.

## 5.3 Resolution change

The change of resolution produced varied differences between the model estimates, whereby total tree number was found to differ by 62355 in 2018, and 81247 in 2021; mean CC by

0.15% in 2018 and 0.98% in 2021 and mean CH by 13.2m in 2018 and 13.3m in 2021. The discrepancies in output can likely be attributed to the distributions of ALS inputs, this is particularly notable in CH, indicating a non random tree distribution, where a greater number of pixels are identified at 20m, decreasing mean CH, as method of calculation was collecting the tallest tree per grid square. Scaling up 20m tree density, N mean estimates to 100m indicates agreement between scales; mean 2018 100m present a mean 20.48 trees per hectare, while upscaled 20m produce 21.2 trees. This is also shown in the 2021 datasets with 19.6 and 20.46 trees for 100m and upscaled 20m respectively. RMSE values also vary between scales, where error was larger for the 20m model for CH and CC by a small margin, however much more noteworthy is the order of magnitude difference between 20m N and 100m N or 0.95 and 9.79 respectively, indicating the increased resolution is superior at understanding tree number in each plot.

Mean estimates of N (scaled), in comparison with mean i-Tree plot data, indicate more similarity with 20m scale outputs, but for mean CC% and CH 100m. However, lower spatial resolution datasets of prediction variables will lead to averaging of responses and often miss out on high resolution variability. Studies with higher resolution tend to be accepted as more "correct" and the coarser resolution datasets attributed to errors (Potapov et al., 2021). Yet through reducing the number of pixels by reducing resolution, can lead to reduced variation and decrease in error as shown in the 100m outputs, albeit limiting the precise identification of specific features. The metrics produced in this study are only estimates and serve as an initial investigation into general distribution across GM, examination of further spatial resolution such as investigating the available 10m Sentinel 2 scale, will likely provide improved clarity into the impact of resolution of RF modelled estimations. Further, control of data collection or sampling of points to ensure the similarity of distributions at varying resolutions would likely also be valuable.

## 5.4 2018 against 2021

This study explains a repeatable framework to monitor change across time. Pixel level outputs from 20m and 100m RF help illustrate that forest structure in GM is dynamic (Figure 14 and 15), demonstrating both positive and negative change spatially throughout GM. Viewing GM holistically however evidenced overall trends between the two years. Total tree number change decreases by ~90,000 using the 20m RF and ~109,000 in the 100m RF, tree losses usually require some level of mitigation planning, as significant time is required to replace

trees and may be replaced with a different size or species, leading to a difference in delivered ecosystem services. The tree loss estimates derived in this study appear to oppose the "ALL OUR TREES" tree and woodland strategy for Greater Manchester collated by City of Trees and Greater Manchester Combined Authority, whose ambition is to deliver net gain across GM, and maximise the benefits from their urban forests. Although, a planting target of three to five million by 2050 exists in line with the Northern Forest project (CITY OF TREES, n.d.), a regional project to plant 50 million trees across northern England; these targets do not account for tree loss. Indeed, in line with national ambitions GM is committed to urban growth providing space for homes and employment. Unfortunately, this can lead to spatial conflicts in greenbelts, around 460 hectares of woodland and 366.5 hectares of trees outside woodlands are located on land allocated for development (CITY OF TREES, n.d.), additionally this value does not include trees already on sites currently in redevelopment. As a rough estimation, using the i-Tree eco estimate of a £4.7 billion replacement value for GMs 11,321,386 trees, produces a replacement value of £415 per tree. Multiplying by the 109,000 tree loss estimate results in a monetary loss of £45.3 million in terms of replacement. While valuation of annual ecosystem service derived from i-Tree Eco surveys of £2.94 per tree, multiplied by loss estimate produces annual benefit losses of £320,594. Although this is only a provisional estimate, as Donovan, (2017) suggests incorporating public health alongside biophysical benefits derived from urban forests, would lead to higher service value and increased investment in maintaining forest assets.

Significant CC change also occurs across GM, with both models agreeing to a net decrease from 2018 to 2021. CC changes are significantly different across RF scales, with a decrease of 19.8 million $m^2$ for the 20m RF and a decrease of 9.8 million $m^2$ for the 100m RF. Although tree planting targets do not account for tree loss, CC targets are independent of tree loss, although GMCA has not set specific CC targets, it aims to ultimately increase this figure. (Doick et al, n.d.) suggests a target for 20% for inland urban areas and 15% for coastal cities, notable as mean CC estimated in 2021 for both 20m and 100m RFs being under 15%.

## 5.5 Future studies

Given the utility of RS in urban ecology, the scope for future studies should not be underestimated. The impact of species on RF models is a vital consideration potentially confounding relationships between variables, and a necessary part of urban biodiversity and GI strategies. The 2018 i-Tree Eco report for GM classified 192 different tree species in plot

data collection, indicating highly significant species heterogeneity ('i-Tree Eco | City of Trees', n.d.). Through utilizing species information, forest strands could be stratified, and help explain species spatial distribution within each borough, and assist in their individual relationship with predictor variables used in this study. Through fusion of hyperspectral and LiDAR, tree species can be accurately modelled for spatially explicit ecosystem assessment (Alonzo et al., 2014), applying species information may lead to reductions in structural variables such as crown canopy size variation.

In addition to RF models trained by ALS, training data extracted from i-Tree Eco plots themselves could be used to train a machine learning model, to then be applied in GM, an approach used in Greater London (Baines et al., 2020). The specific heterogenous nature of UK cities could train i-Tree RF models and then tested against each other to elicit the most appropriate models for use in other urban areas, where LiDAR or plot data collection is limited. Forest structure results should be utilized in larger scale synoptic urban analysis, in combination with other features of the urban matrix such as proximity to population centres, or grey infrastructure to assist in explaining the relationship between urban forests and derived ecosystem services.

This methodological framework should also be applied to different multispectral predictor variable datasets of higher spatial resolution such as the PlanetScope constellation, with resolutions of 3m to better determine how improved spatial resolutions can impact accuracy of forest inventory. Further, although the temporal aspect of change detection was briefly inspected in this study, only two data points, 2018 and 2021 were examined. Long term datasets such as the LandSat programme would assess the longitudinal variation in forest structure over decadal time frames, or utlising the relatively rapid return time of Sentinel 2 constellation to examine phenological variation, extracting trends and producing data to better inform long term sustainable urban forest management.

## 5.6 Limitations

The pursual of this research project, utilizing remote sensing in an ecological context was driven through an interest in how earth observation technologies can be used to meet global challenges. This project combining passive and active remote sensing, and machine learning in an open and free methodology was intended to be novel and useful for several stakeholders and decisionmakers. This open approach facilitates the uptake of both method and results by land and city managers. The project originally planned to be executed through GEE, utilizing the benefits of the cloud platform; however, memory constraints led to training and applying RF models natively in R. This constraint was not an issue in the context of an academic research project, allowing exploration and creation of novel datasets, building skills in multiple programming languages and develop an understanding of machine learning. However, in an operational context where resources may be limited and decision making cycles are accelerated, difficulties may arise where non expert users cannot fully explain results or processes and "debug" software. Nevertheless, this open approach simplifies traditional modelling approaches through using analysis ready data, preprocessed in GEE, and lowering the technical knowledge required to undertake analyses.

The mapping of forest structure across a large metropolitan area such as GM, is significant to monitoring urban forests, particularly given the high revisit frequency of Sentinel 2 constellation, and 20m resolution models, producing medium spatial and high temporal resolution leading to identification of patterns and trends of forest structure with improved granularity. Further, the derivation of key structure metrics such as canopy height and canopy cover, can facilitate the study of other metrics such as aboveground biomass, through the previously established relationships between variables.

However, multiple limitations were identified in the project. Firstly, the accuracies canopy height and tree numbers were appreciably lower than other urban forest studies; suggesting, the application of results should be questioned before used. Input data was not resampled to normal distributions, and so results may be less accurate.

EA LiDAR point cloud pulse density is coarse (1-2 $m^{-1}$) significantly impacting data capture, potentially leading to many small trees unaccounted and lower tree number estimates. Although known for being expensive, it may be pertinent to acquire high pulse density LiDAR data through ALS, TLS or MLS to construct improved accuracy in training datasets and

therefore forest structure models and metric estimates. Additionally, the relatively small area of ALS surveys, only 3.9% of GM, undertaken in only one geographical region may have hindered result accuracy due to correlated land typologies. If the project were to be undertaken again, efforts to collect LiDAR data across GM would likely produce more representative training datasets, in addition to applying other tree delineation algorithms, as only one was used in this study, potentially leading to missed tree data.

Although i-Tree Eco measurement error for in the field data collection, is generally assumed to zero (Alonzo et al., 2016), exploration of plot level and observed tree datasets discovered discrepancies. In some plots where canopy cover was identified, the related tree metrics for that plot displayed N/A values indicating either data was not collected or a discrepancy was found. This severely limits the use of GM i-Tree Eco datasets for further analysis, as the true value cannot be ascertained. However, even accounting for the limitations described, this project contributes to the field of urban ecology and is intended to assist land managers with monitoring and promoting sustainable urban strategies.

# 6. Conclusion

Continued monitoring of urban forests is vital to assist in quantification of ecosystem services and how urban forest structure change through time. This process is required for the maintenance of sustainable green infrastructure assets, and to maximise benefits to urban residents. This analysis in this study produced wall to wall maps of three forest structure metrics: canopy height, canopy cover and number of trees represented at 20m and 100m resolutions, over two time periods and across a spatially large and heterogenous study area of GM. Mapping was derived using an open, flexible and modular methodology. Active LiDAR, passive Sentinel 2 and supplementary datasets were fused in an RF model. In total using our 20m model, an estimated 2.7 million trees were discovered across GM in the year 2018; with spatial data indicating a non-uniform distribution across GM boroughs. Average canopy cover in 2018 was found to be 14.8%, whereas average tree height was 5.6m. Through changing resolution of predictor variables illustrated changes in estimated outputs, however both resolutions were found to underestimate total number of trees compared to i-Tree Eco reports. Changes in urban forest structure between the years of 2018 and 2021 indicate a net decrease in total number of trees and canopy cover in GM, although forest structure change across local authority districts is not equal, where some had net increases. The results from this analysis have the potential to facilitate finer scale analyses across large areas. The method using RS techniques provides critical forestry information for urban land managers for the improved monitoring and maintenance of urban forests, and the ecosystem services they afford. Given the open source nature of this approach, it is suggested the method could be applied in other metropolitan regions where data is sparse, or data collection not practical. Additionally, these methods could be relatively simply and inexpensively adopted by poorer nations worldwide, assisting in the safeguarding of global urban forests. Through the improved quantification of urban forest structure in the UK and beyond, cities can be steered to meet the objectives of UN SDG 11, through mitigating and adapting to the adverse effects of combined population growth and climatic change pressures.

## Appendices

## References


Ahmed, O.S., Franklin, S.E., Wulder, M.A. & White, J.C. (2015) 'Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm', *ISPRS Journal of Photogrammetry and Remote Sensing*, 101pp. 89–101.

Alonzo, M., Bookhagen, B. & Roberts, D.A. (2014) 'Urban tree species mapping using hyperspectral and lidar data fusion', *Remote Sensing of Environment*, 148pp. 70–83.

Alonzo, M., McFadden, J.P., Nowak, D.J. & Roberts, D.A. (2016) 'Mapping urban forest structure and function using hyperspectral imagery and lidar data', *Urban Forestry & Urban Greening*, 17pp. 135–147.

Amani, M., Ghorbanian, A., Ahmadi, S.A., Kakooei, M., Moghimi, A., Mirmazloumi, S.M., Moghaddam, S.H.A., Mahdavi, S., Ghahremanloo, M., Parsian, S., Wu, Q. & Brisco, B. (2020) 'Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13pp. 5326–5350.

Anon (2018) *i-Tree Eco | City of Trees*. [Online] [online]. Available from: https://www.cityoftrees.org.uk/project/i-tree-eco (Accessed 1 August 2022).

Anon (n.d.) *i-Tree Eco | City of Trees*. [Online] [online]. Available from: https://www.cityoftrees.org.uk/project/i-tree-eco (Accessed 9 March 2022).

Anon (2022) *i-Tree Eco | i-Tree*. [Online] [online]. Available from: https://www.itreetools.org/tools/i-tree-eco (Accessed 6 April 2022).

Baines, O., Wilkes, P. & Disney, M. (2020) 'Quantifying urban forest structure with open-access remote sensing data sets', *Urban Forestry & Urban Greening*, 50p. 126653.

BLUESKY (2017) *BLUESKY National Tree Map (NTM) - Catchment Management Modelling Platform*. [Online] [online]. Available from: https://catalogue.ceh.ac.uk/documents/0bdac1d1-3c3f-4538-b35d-ffb3d8dcd0ab (Accessed 9 April 2022).

Bolund, P. & Hunhammar, S. (1999) 'Ecosystem services in urban areas', *Ecological Economics*, 29(2), pp. 293–301.

Boyd, D.S., Foody, G.M., Brown, C., Mazumdar, S., Marshall, H. & Wardlaw, J. (2022) 'Citizen science for Earth Observation (Citizens4EO): understanding current use in the UK', *https://doi.org/10.1080/01431161.2022.2076574*, 43(8), pp. 2965–2985.

Boyd, J. & Banzhaf, S. (2007) 'What are ecosystem services? The need for standardized environmental accounting units', *Ecological Economics*, 63(2–3), pp. 616–626.

Breiman, L. (2001) 'Random Forests', *Machine Learning 2001 45:1*, 45(1), pp. 5–32.

Bruggisser, M., Hollaus, M., Kükenbrink, D. & Pfeifer, N. (2019) 'COMPARISON of FOREST STRUCTURE METRICS DERIVED from UAV LIDAR and ALS DATA', *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(2/W5), pp. 325–332.

Cameron, R.W.F. & Blanuša, T. (2016a) 'Green infrastructure and ecosystem services – is the devil in the detail?', *Annals of Botany*, 118(3), pp. 377–391.

Cameron, R.W.F. & Blanuša, T. (2016b) 'Green infrastructure and ecosystem services – is the devil in the detail?', *Annals of Botany*, 118(3), pp. 377–391.

Centre for Cities (2021) *Greater Manchester* . [Online] [online]. Available from: https://www.centreforcities.org/combined-authority/greater-manchester/ (Accessed 7 April 2022).

Chrysoulakis, N., Somarakis, G., Stagakis, S., Mitraka, Z., Wong, M.-S. & Ho, H.-C. (2021) 'Monitoring and Evaluating Nature-Based Solutions Implementation in Urban Areas by Means of Earth Observation', *Remote Sensing 2021, Vol. 13, Page 1503*, 13(8), p. 1503.

CITY OF TREES (n.d.) *All Our Trees: Greater Manchester's Tree & Woodland Strategy | City of Trees*. [Online] [online]. Available from: https://www.cityoftrees.org.uk/allourtrees (Accessed 5 July 2022).

Delegido, J., Verrelst, J., Alonso, L. & Moreno, J. (2011) 'Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content', *Sensors (Basel, Switzerland)*, 11(7), p. 7063.

Doick et al (n.d.) *The Canopy Cover of England's Towns and Cities (Research paper) - The Institute of Chartered Foresters*. [Online] [online]. Available from: https://www.charteredforesters.org/resource/doick-et-al-the-canopy-cover-of-englands-towns-and-cities-research-paper (Accessed 6 March 2022).

Doick, K., Davies, H., Handley, P., Vaz Monteiro, M., O'Brien, L. & Ashwood, F. (2016) *Introducing England's Urban Forests*.

Doick, K., Handley, P., Davies, H., O'Brien, L. & Wilson, J. (2017) *Delivery of ecosystem services by urban forests - Forest Research*.

Doick, K.J., Davies, H.J., Moss, J., Coventry, R., Handley, P., Vazmonteiro, M., Rogers, K. & Simpkin, P. (n.d.) *The Canopy Cover of England's Towns and Cities: baselining and setting targets to improve human health and well-being*,

Donager, J.J., Sánchez Meador, A.J., Blackburn, R.C., Donager, C.:, Meador, S. & Blackburn, A.J.; (2021) 'Comparison of forest structure metrics derived from UAV lidar and ALS data', *pdfs.semanticscholar.org*,

Donovan, G.H. (2017) 'Including public-health benefits of trees in urban-forestry decision making', *Urban Forestry & Urban Greening*, 22pp. 120–123.

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W. & Li, X. (2016) 'Water Bodies' Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution Produced by Sharpening the SWIR Band', *Remote Sensing 2016, Vol. 8, Page 354*, 8(4), p. 354.

Duan, Q., Tan, M., Guo, Y., Wang, X. & Xin, L. (2019a) 'Understanding the Spatial Distribution of Urban Forests in China Using Sentinel-2 Images with Google Earth Engine', *Forests 2019, Vol. 10, Page 729*, 10(9), p. 729.

Duan, Q., Tan, M., Guo, Y., Wang, X. & Xin, L. (2019b) 'Understanding the Spatial Distribution of Urban Forests in China Using Sentinel-2 Images with Google Earth Engine', *Forests 2019, Vol. 10, Page 729*, 10(9), p. 729.

Endreny, T., Santagata, R., Perna, A., … C.D.S.-E. & 2017, undefined (2017) 'Implementing and managing urban forests: A much needed conservation strategy to increase ecosystem services and urban wellbeing', *Elsevier*,

Endreny, T.A. (2018) 'Strategically growing the urban forest will improve our world', *Nature Communications 2018 9:1*, 9(1), pp. 1–3.

Escobedo, F.J., Kroeger, T. & Wagner, J.E. (2011) 'Urban forests and pollution mitigation: analyzing ecosystem services and disservices', *Environmental pollution (Barking, Essex : 1987)*, 159(8–9), pp. 2078–2087.

FAO (2016) *Guidelines on urban and peri-urban forestry*.

Fisher, B. & Kerry Turner, R. (2008) 'Ecosystem services: Classification for valuation', *Biological Conservation*, 141(5), pp. 1167–1169.

Forest Research (2017a) *i-Tree Eco Edinburgh - Forest Research*.

Forest Research (2015a) *i-Tree Eco Glasgow - Forest Research*.

Forest Research (2015b) *i-Tree Eco London - Forest Research*.

Forest Research (2020) *i-Tree Eco Newport - Forest Research*.

Forest Research (2017b) *i-Tree Eco Petersfield - Forest Research*.

Forest Research (2017c) *i-Tree Eco Southampton - Forest Research*.

Gao, J. & O'Neill, B.C. (2020) 'Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways', *Nature Communications 2020 11:1*, 11(1), pp. 1–12.

GMCA (2019) *Five-Year Environment Plan - Greater Manchester Combined Authority*. [Online] [online]. Available from: https://www.greatermanchester-ca.gov.uk/what-we-do/environment/five-year-environment-plan/ (Accessed 5 July 2022).

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. & Moore, R. (2017) 'Google Earth Engine: Planetary-scale geospatial analysis for everyone', *Remote Sensing of Environment*, 202pp. 18–27.

Hansen, M.C., Potapov, P. v., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S. v., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O. & Townshend, J.R.G. (2013) 'High-resolution global maps of 21st-century forest cover change', *Science*, 342(6160), pp. 850–853.

Hanssen, F., Barton, D.N., Venter, Z.S., Nowell, M.S. & Cimburova, Z. (2021) 'Utilizing LiDAR data to map tree canopy for urban ecosystem extent and condition accounts in Oslo', *Ecological Indicators*, 130p. 108007.

Hassaan, O., Nasir, A.K., Roth, H. & Khan, M.F. (2016) 'Precision Forestry: Trees Counting in Urban Areas Using Visible Imagery based on an Unmanned Aerial Vehicle', *IFAC-PapersOnLine*, 49(16), pp. 16–21.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) 'Very high resolution interpolated climate surfaces for global land areas', *International Journal of Climatology*, 25(15), pp. 1965–1978.

Hird, J.N., DeLancey, E.R., McDermid, G.J. & Kariyeva, J. (2017) 'Google Earth Engine, Open-Access Satellite Data, and Machine Learning in Support of Large-Area Probabilistic Wetland Mapping', *Remote Sensing 2017, Vol. 9, Page 1315*, 9(12), p. 1315.

Huang, C., Zhang, C., He, Y., Liu, Q., Li, H., Su, F., Liu, G. & Bridhikitti, A. (2020) 'Land Cover Mapping in Cloud-Prone Tropical Areas Using Sentinel-2 Data: Integrating Spectral Features with Ndvi Temporal Dynamics', *Remote Sensing 2020, Vol. 12, Page 1163*, 12(7), p. 1163.

Kuhn, M., Wing, J., Weston, S., Williams, A.W., Keefer, C. & Engelhardt, A. (n.d.) *CRAN - Package caret*.

Lang, N., Schindler, K. & Wegner, J.D. (2019) 'Country-wide high-resolution vegetation height mapping with Sentinel-2', *Remote Sensing of Environment*, 233p. 111347.

Lechner, A.M., Foody, G.M. & Boyd, D.S. (2020) 'Applications in Remote Sensing to Forest Ecology and Management', *One Earth*, 2(5), pp. 405–412.

Li, X., Chen, W.Y., Sanesi, G. & Lafortezza, R. (2019) 'Remote Sensing in Urban Forestry: Recent Applications and Future Directions', *Remote Sensing 2019, Vol. 11, Page 1144*, 11(10), p. 1144.

Liaw, A. & Wiener, M. (2002) *Classification and Regression by randomForest*, 2(3), .

Lin, J., Kroll, C.N. & Nowak, D.J. (2021) 'An uncertainty framework for i-Tree eco: A comparative study of 15 cities across the United States', *Urban Forestry & Urban Greening*, 60p. 127062.

Matasci, G., Coops, N.C., Williams, D.A.R. & Page, N. (2018) 'Mapping tree canopies in urban environments using airborne laser scanning (ALS): A vancouver case study', *Forest Ecosystems*, 5(1), pp. 1–9.

Mather, P.M. (2011) *Computer processing of remotely sensed images : an introduction*. 4th edition. Vol. 5.

Mellor, A., Haywood, A., Stone, C. & Jones, S. (2013) 'The Performance of Random Forests in an Operational Setting for Large Area Sclerophyll Forest Classification', *Remote Sensing 2013, Vol. 5, Pages 2838-2856*, 5(6), pp. 2838–2856.

Monteiro, M.V., Handley, P. & Doick, K.J. (2020) 'An insight to the current state and sustainability of urban forests across Great Britain based on i-Tree Eco surveys', *Forestry: An International Journal of Forest Research*, 93(1), pp. 107–123.

Mutanga, O. & Kumar, L. (2019) 'Google Earth Engine Applications', *Remote Sensing 2019, Vol. 11, Page 591*, 11(5), p. 591.

Natural England (2009) *Natural England's Green Infrastructure Guidance - NE176*.

Nesbitt, L., Hotte, N., Barron, S., … J.C.-U.F.& U. & 2017, undefined (2017) 'The social and economic value of cultural ecosystem services provided by urban forests in North America: A review and suggestions for future research', *Elsevier*,

Neuville, R., Bates, J.S. & Jonard, F. (2021) 'Estimating Forest Structure from UAV-Mounted LiDAR Point Cloud Using Machine Learning', *Remote Sensing 2021, Vol. 13, Page 352*, 13(3), p. 352.

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S. & Crowston, K. (2012) 'The future of citizen science: emerging technologies and shifting paradigms', *Frontiers in Ecology and the Environment*, 10(6), pp. 298–304.

Nitoslawski, S.A., Galle, N.J., van den Bosc, C.K. & Steenberg, J.W.N. (2019) 'Smarter ecosystems for smarter cities? A review of trends, technologies, and turning points for smart urban forestry', *Sustainable Cities and Society*, 51p. 101770.

ONS (2021) *Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics*. [Online] [online]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland (Accessed 7 April 2022).

Parmehr, E.G., Amati, M. & Fraser, C.S. (2016) 'MAPPING URBAN TREE CANOPY COVER USING FUSED AIRBORNE LIDAR and SATELLITE IMAGERY DATA', *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3pp. 181–186.

Patela, N.N., Angiuli, E., Gamba, P., Gaughan, A., Lisini, G., Stevens, F.R., Tatem, A.J. & Trianni, G. (2015) 'Multitemporal settlement and population mapping from Landsat using Google Earth Engine', *International Journal of Applied Earth Observation and Geoinformation*, 35(PB), pp. 199–208.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C.E., Armston, J., Dubayah, R., Blair, J.B.

& Hofton, M. (2021) 'Mapping global forest canopy height through integration of GEDI and Landsat data', *Remote Sensing of Environment*, 253p. 112165.

Puissant, A., Rougiera, S. & Stumpf, A. (2014) 'Object-oriented mapping of urban trees using Random Forest classifiers', *International Journal of Applied Earth Observation and Geoinformation*, 26(1), pp. 235–245.

QGIS Development Team (2022) *QGIS Geographic Information System' Open Source Geospatial Foundation Project [Online]*.

R Core Team (2018) *R: The R Project for Statistical Computing*.

Raum, S., Hand, K.L., Hall, C., Edwards, D.M., O'Brien, L. & Doick, K.J. (2019) 'Achieving impact from ecosystem assessment and valuation of urban greenspace: The case of i-Tree Eco in Great Britain', *Landscape and Urban Planning*, 190p. 103590.

Roman, L.A., Scharenbroch, B.C., Östberg, J.P.A., Mueller, L.S., Henning, J.G., Koeser, A.K., Sanders, J.R., Betz, D.R. & Jordan, R.C. (2017) 'Data quality in citizen science urban tree inventories', *Urban Forestry & Urban Greening*, 22pp. 124–135.

Romolini, M., Brinkley, W. & Wolf, K.L. (2012) *What is urban environmental stewardship? Constructing a practitioner-derived framework*,

Roussel, J.-R. & Auty, D. (2021) *LidR - Airborne LiDAR Data Manipulation and Visualization for Forestry Applications.* .

Roy, S., Byrne, J. & Pickering, C. (2012) 'A systematic quantitative review of urban tree benefits, costs, and assessment methods across cities in different climatic zones', *Urban Forestry & Urban Greening*, 11(4), pp. 351–363.

Ruiz, I., Elias Ruiz Hernandez, I., Shi, W. & Ruiz Hernandez, I.E. (2017) 'A Random Forests classification method for urban land-use mapping integrating spatial metrics and texture analys... International Journal of Remote Sensing A Random Forests classification method for urban land-use mapping integrating spatial metrics and texture analysis A Random Forests classification method for urban land-use mapping

integrating spatial metrics and texture analysis', *International Journal of Remote Sensing*, 39(4), pp. 1175–1198.

Salmond, J.A., Tadaki, M., Vardoulakis, S., Arbuthnott, K., Coutts, A., Demuzere, M., Dirks, K.N., Heaviside, C., Lim, S., MacIntyre, H., McInnes, R.N. & Wheeler, B.W. (2016) 'Health and climate related ecosystem services provided by street trees in the urban environment', *Environmental Health: A Global Access Science Source*, 15(1), pp. 95–111.

Seddon, N. (2022) 'Harnessing the potential of nature-based solutions for mitigating and adapting to climate change', *Science*, 376(6600), pp. 1410–1416.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S. & Brisco, B. (2020) 'Google Earth Engine for geo-big data applications: A meta-analysis and systematic review', *ISPRS Journal of Photogrammetry and Remote Sensing*, 164pp. 152–170.

UN - HABITAT (n.d.) *New Urban Agenda | UN-Habitat*. [Online] [online]. Available from: https://unhabitat.org/about-us/new-urban-agenda (Accessed 1 August 2022).

UN Department of Economic and Social Affairs (2018) *2018 Revision of World Urbanization Prospects | Multimedia Library - United Nations Department of Economic and Social Affairs*. [Online] [online]. Available from: https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html (Accessed 10 March 2022).

UN HABITAT (2011) *Global Report on Human Settlements 2011: Cities and Climate Change | UN-Habitat*.

Wallner, A., Elatawneh, A., Schneider, T. & Knoke, T. (2015) 'Estimation of forest structural information using RapidEye satellite data', *Forestry: An International Journal of Forest Research*, 88(1), pp. 96–107.

Walters, M. & Sinnett, D. (2021) 'Achieving tree canopy cover targets: A case study of Bristol, UK', *Urban Forestry & Urban Greening*, 65p. 127296.

Wang, K., Franklin, S.E., Guo, X. & Cattet, M. (2010) 'Remote Sensing of Ecology, Biodiversity and Conservation: A Review from the Perspective of Remote Sensing Specialists', *Sensors 2010, Vol. 10, Pages 9647-9667*, 10(11), pp. 9647–9667.

Weinstein, B.G., Marconi, S., Bohlman, S., Zare, A. & White, E. (2019) 'Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks', *Remote Sensing 2019, Vol. 11, Page 1309*, 11(11), p. 1309.

Wilkes, P., Disney, M., Vicari, M.B., Calders, K. & Burt, A. (2018) 'Estimating urban above ground biomass with multi-scale LiDAR', *Carbon Balance and Management*, 13(1), pp. 1–20.

Wilkes, P., Jones, S.D., Suarez, L., Mellor, A., Woodgate, W., Soto-Berelov, M., Haywood, A. & Skidmore, A.K. (2015) 'Mapping Forest Canopy Height Across Large Areas by Upscaling ALS Estimates with Freely Available Satellite Data', *Remote Sensing 2015, Vol. 7, Pages 12563-12587*, 7(9), pp. 12563–12587.

Woodland Trust (2022) *The Northern Forest - .* [Online] [online]. Available from: https://www.woodlandtrust.org.uk/about-us/what-we-do/we-plant-trees/the-northern-forest/ (Accessed 7 April 2022).

Xu, H. (2007) 'Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery', *https://doi.org/10.1080/01431160600589179*, 27(14), pp. 3025–3033.

Zhang, C., Zhou, Y. & Qiu, F. (2015) 'Individual Tree Segmentation from LiDAR Point Clouds for Urban Forest Inventory', *Remote Sensing 2015, Vol. 7, Pages 7892-7913*, 7(6), pp. 7892–7913.

Zhu, Z., Zhou, Y., Seto, K.C., Stokes, E.C., Deng, C., Pickett, S.T.A. & Taubenböck, H. (2019) 'Understanding an urbanizing planet: Strategic directions for remote sensing', *Remote Sensing of Environment*, 228pp. 164–182.