GLA INTELLIGENCE

Briefing 2

How Green is London?

GLA City Intelligence

January 2019

MAYOR OF LONDON

copyright

Greater London Authority
January 2019

Published by
Greater London Authority
City Hall
The Queens Walk
London SE1 2AA

www.london.gov.uk
Tel 020 7983 4000

Minicom 020 7983 4458

Cover photograph © By Nadiia Gerbish | Shutterstock

For more information about this publication, please contact:

GLA Intelligence

Tel 020 7983 5742

Email gis@london.gov.uk

Data and analysis from GLA City Intelligence form a basis for the policy and investment decisions facing the Mayor of London and the GLA group. GLA City Intelligence uses a wide range of information and data sourced from third party suppliers within its analysis and reports. GLA City Intelligence cannot be held responsible for the accuracy or timeliness of this information and data.

The GLA will not be liable for any losses suffered or liabilities incurred by a party as a result of that party relying in any way on the information contained in this report.

Contents

Executive summary

The GLA have produced a new dataset that provides a more accurate estimate of the extent of the London's green infrastructure - the city's parks, gardens, trees, green spaces, rivers and wetlands, and features such as green roofs.

Previous assessments have been based on less accurate data and have produced aggregated estimates for geographies such as London Boroughs. The analysis presented here creates a detailed high-resolution green cover layer for the Greater London area. By combining classified near-infrared aerial imagery (NDVI) with land use datasets, small areas such as private gardens and tree canopy can be captured in a way that was not possible with previous methodologies.

The results of the analysis indicates that London's green cover is between 48-51 percent. The baseline is presented as a range to account for variations in the analysis of aerial imagery. The following report sets out the methodology behind the creation of the new map.

Although not without its limitations from issues such as: image shadow, unhealthy vegetation and land use uncertainties, the approach taken is robust and produces a layer that performs well in an accuracy assessment (96.5% overall accuracy). The methodology has been published in a transparent and open manner, along with processing scripts and the final spatial layer created. The intention is for this to allow further analysis and refinement of the approach by other interested parties.

Introduction

All cities now recognise that the green infrastructure of parks, green spaces, natural habitats, street trees, green roofs and walls are an essential part of city life. This green infrastructure helps to make the city healthy, liveable, resilient to climate change, and more economically sustainable.

In the face of the additional development required to meet the demands of a growing population, there is a need to significantly increase the area of green cover in the built environment, whilst maintaining and enhancing the existing network of parks and green spaces to ensure that London becomes greener overall.

In the London Environment Strategy published May 2018, the Mayor committed to make more than half of London green by 2050. To be able to do track progress on this, as well as target areas for further greening, a complete picture of London's green infrastructure was needed. To do this, existing information and London's green and blue cover was reviewed.

Green cover represents the total area covered by vegetation and water across London. It not only includes publicly accessible and publicly managed vegetated land (i.e. green space), but also non-accessible green and blue spaces, as well as privately owned vegetated land including farmland, private gardens as well as the area of vegetated cover on buildings and in the wider built environment such as green roofs, street trees and rain gardens. Blue cover refers to the subset of this that is predominately water.

Why was the map created?

Previously, the best available data could not be used to monitor London's progress against this target as it did not map green and blue cover in sufficient detail. In particular, previous analysis did not capture private green and blue spaces, isolated green areas, such as grass verges, or other green infrastructure, such as green roofs or isolated trees in a spatial way. This new map provides this and acts as the baseline for measuring future change.

Data Sources and Processing

A number of data sources, including open datasets, were evaluated in the research phase of the project. Some open source satellite imagery had appropriate spatial coverage and temporal resolution but the spatial resolution captured was not detailed enough to identify green cover at the level required accurately. The Ordnance Survey MasterMap Greenspace layer is potentially useful but did not include the isolated green cover, such as street trees, needed for this project.

After careful consideration, a combination of the following datasets were identified as most suitable:

1. Aerial imagery: high quality pan-London imagery captured during September 2016.

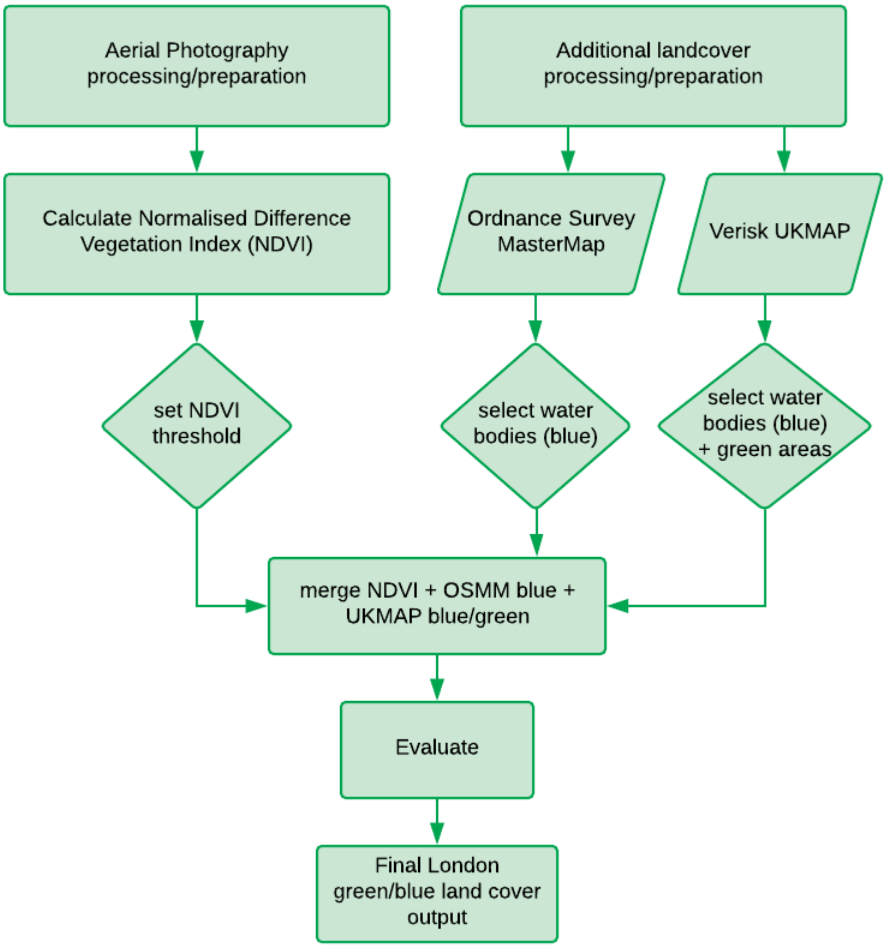
2. Ordnance Survey MasterMap: topography layer representing real-world features as points, lines and polygons.

3. Verisk GeoInformation Group UKMap: topographic map of polygons classified by the type of feature e.g. building, man-made surface…etc

In summary, these datasets were analysed, processed and combined to generate a single layer of green and blue cover for London. As discussed in this document, additional land cover datasets (MasterMap and UKMap) were used to help represent green/blue land cover that is otherwise obscured in the imagery data. For instance, to fill in temporary unhealthy bald patches in the middle of playing fields.

Figure 1 provides a simplified overview of the process which is described in more detail in the Methodology section of this document.

Figure 1: Overview of how the datasets were combined to a green and blue cover layer for London.



Source: GLA GIS Team

More details of these three datasets are described below.

1. **Aerial Imagery** (Licensed dataset)

The key dataset within this project was the high-resolution aerial imagery captured in September 2016 (Table 1). Each pixel covered 10cm of the ground surface, and so, in most cases, would easily be able to distinguish small areas of green cover. The imagery was captured with 4 wavelength bands: red, green, blue and near infrared. Displaying the image bands red, green and blue presents the images as a true colour and is how the human eye would see the image i.e. like a traditional photograph. Choosing to display the near infrared band with the green and blue channel will generate a color infrared image (Map 1). This near-infrared image is used to assess vegetation cover, see the *Methodology* section for more details.

Table 1: aerial imagery characteristics

Spatial resolution/scale	25cm (resampled from 10cm)
Acquisition date	September 2016
Data provider / licensor	Verisk GeoInformation
Original format	ECW
Format for data processing	GeoTiff
Coverage	London (chunked into 467 tiles at ~2km x 2km)
Bands	Band 1: Red channel: Near Infrared band Band 2: Green channel: Red band Band 3: Blue channel: Green band

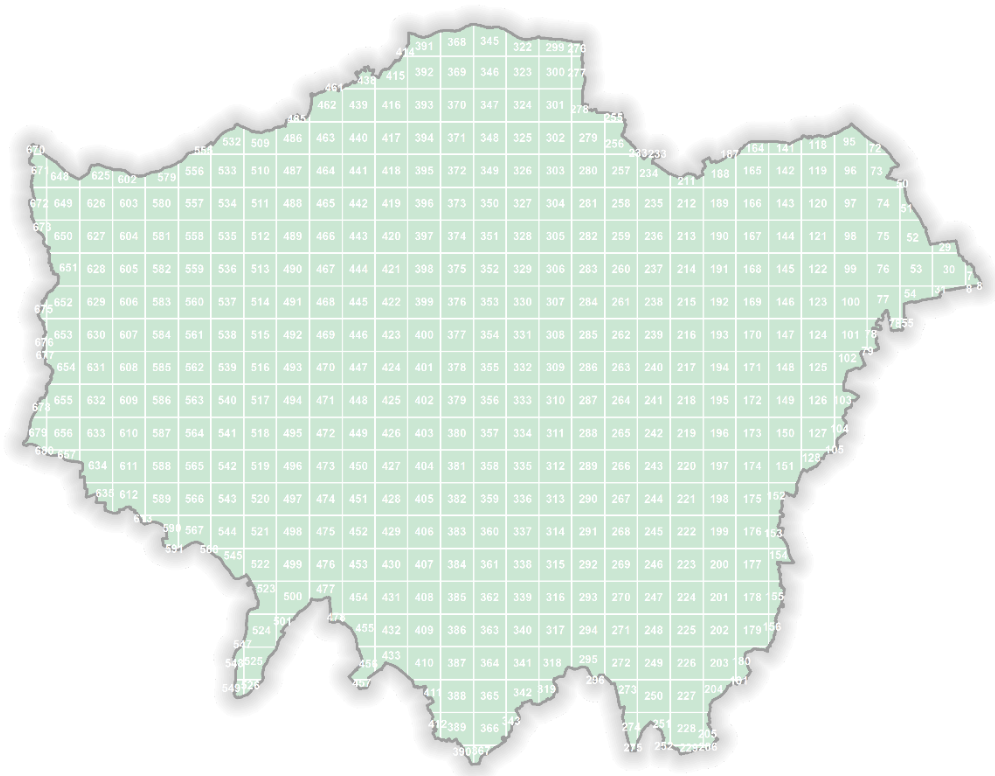
Source: Verisk Analytics GeoInformation 2016

Map 1. Aerial imagery wavelength bands that generate both true colour and near infrared images

Originally supplied in the highly compressed ECW format, the uncompressed data expands to several hundred Gigabits. Owing to concerns on data volume and data processing time, it was decided that resampling the imagery to reduce the number of pixels would be a sensible approach. Following the testing of different resampling levels with complex areas, such as allotments, 25cm was deemed most suitable, as, it was still possible to distinguish small patches of green cover.

The single ECW file also needed to be tiled into multiple smaller files for more efficient processing. Various tile sizes were considered, with 2km by 2km size tiles chosen as an ideal size in terms of processing time and for sharing outputs. A 2km by 2km tile grid covering the full extent of the Greater London boundary (within the GLA Boundary) was generated, containing 467 tiles in the GeoTIFF format (Map 2).

Map 2. The 2km by 2km tile grid used for the data processing



2. **Ordnance Survey MasterMap** (Licensed dataset)

In addition to green cover, blue cover was also estimated using a combination of the OS MasterMap and UKMap landuse datasets. It was decided to use both sources as each dataset contained waterbodies that were not present in the other. The OS MasterMap Topography layer characterises features as points, lines or polygons and groups these into themes such as: buildings, rail, water, etc. The data is captured by a combination of ground and aerial surveys.

Table 2: OS MasterMap characteristics

Spatial resolution/scale	1:1250
Acquisition date	July 2018
Data provider / licensor	Ordnance Survey
Original format	GML (Geographic Markup Language)
Format for data processing	SHP/GDB (native ESRI software formats)
Coverage	London

Source: Ordnance Survey

Water bodies or 'blue cover' were extracted from this layer using the query:

Theme = 'water'

Some duplicates were found within the initial selection, but consequently dealt with by dissolving on the TOID field – a unique identifier. Total blue area was calculated at 43,758,763.09 m², which is approximately 2.7% of London land extent (Map 3).

Map 3: Ordnance Survey MasterMap water body selection

3. **Verisk GeoInformation Group UKMap** (Licensed dataset)

The UKMap product features a polygon Base layer that classifies each polygon with a land use code, the Feature Classification Code (FCC). This is based on data collected in the field or interpreted from the aerial photography and provides a detailed classification of land use types based on the National Land Use Specification Database (NLUD).

Table 3: GeoInformation UKMap characteristics

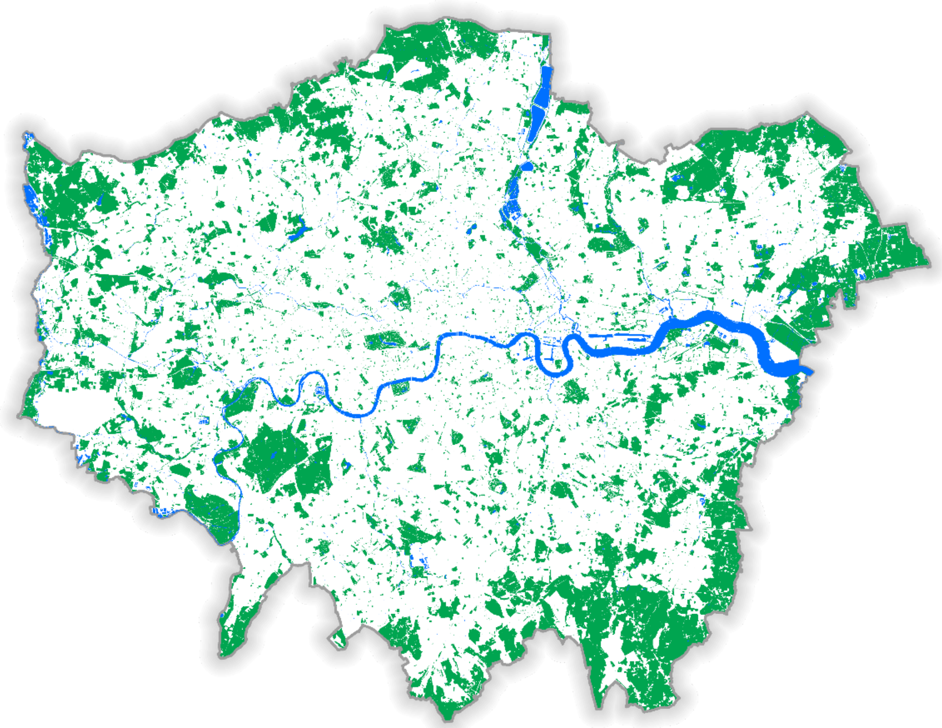
Spatial resolution/scale	1:1250
Acquisition date	October 2017
Data provider / licensor	Verisk Analytics GeoInformation Group
Original format	GDB (geodatabase, native ESRI format)
Format for data processing	SHP/GDB (native ESRI software formats)
Coverage	London

Source: Verisk Analytics GeoInformation Group

In addition to the FCC, there is a primary classification called Feature Type Code (FTC), which is a more generic high-level classification such as building, vegetation or man-made surface. An initial evaluation of these codes created a selection of FCC/FTC codes that could be described as representative of green and blue land cover. These codes and descriptions are provided in Appendix I.

A selection of relevant FCC and FTC codes produces a blue area of 42,369,968.98 m² and green of 412,194,773.56 m², this represents 2.7% and 25.9% of London's land extent (Map 4). While this provides a good base for calculating green and blue cover, it does not include the majority of private, small and isolated green infrastructure that make up part of the complete picture.

Map 4: UKMap blue and green cover selection



Technology / Software

Various software approaches were tried and tested during early stages of this project. Following this discovery phase, it was decided to adopt a primarily open source approach using the *Python* programming language where possible. This would enable the project to be repeatable in subsequent years at little extra cost. The methodology and code will be easily and freely accessible to others wanting to complete similar research. Whilst open Python packages were used as far as possible, some proprietary software was used, but it would be hoped that with more time and expertise, the processes for which these were used would be replaced with open source alternatives.

Key Python packages used in this project included GDAL, NumPy, Pandas and GeoPandas. These were managed in a *Conda* environment, using Python 3.6. The majority of image processing and data analysis was completed using Python.

Additionally, the ESRI ArcGIS ArcPy package was also used for dissolve and clip functions. This was due to some clip and dissolve processes frequently failing whilst using GeoPandas.

SAFE *FME* software was used for reading, resampling and tiling the ECW aerial image file, as well as processing the landcover datasets.

Data processing scripts are available on the GLA GitHub site.

Finally, for sharing data, a webmap was developed using ESRI ArcGIS Server and ESRI Leaflet JavaScript API.

Links to the webmap and processing scripts can be found on the London Datastore website.

Methodology

This section covers each step of the methodology in detail. Figure 2, at the end of this section, presents a detailed flow diagram of each step.

1. Data preparation

As discussed in the *Data Sources and Processing* section of this document, high resolution aerial imagery was resampled to 25cm pixel size, and tiled into 467 2km by 2km tiles in GeoTIFF format.

Since this project aimed to produce a comprehensive green and blue cover map, additional landcover datasets were also required for two main reasons. Firstly, the image analysis would only capture healthy green vegetation that was unaffected by shadow and other obstructions. Secondly, blue cover was often hidden by tree canopy and other features, therefore it would be very difficult to extract this from the aerial imagery.

2. Normalised Difference Vegetation Index

The Normalised Difference Vegetation Index (NDVI) is an effective, widely used, measure for determining the presence of healthy green vegetation. Chlorophyll contained within plants absorb visible light but reflects near-infrared (NIR) light. By applying the NDVI formula (Fig X) to the aerial imagery red and NIR bands, a value between -1 and 1 is calculated for each pixel.

Fig X: Normalised Difference Vegetation Index formula

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

Where: NIR = Near-Infrared radiation, Red = red radiation band

Source: NASA 2000.

Values closer to 1 indicates presence of healthy vegetation; as values approach 0, less vegetation is present.

The NDVI was calculated at the 25cm pixel resolution for each aerial imagery tile using the Python GDAL and NumPy packages.

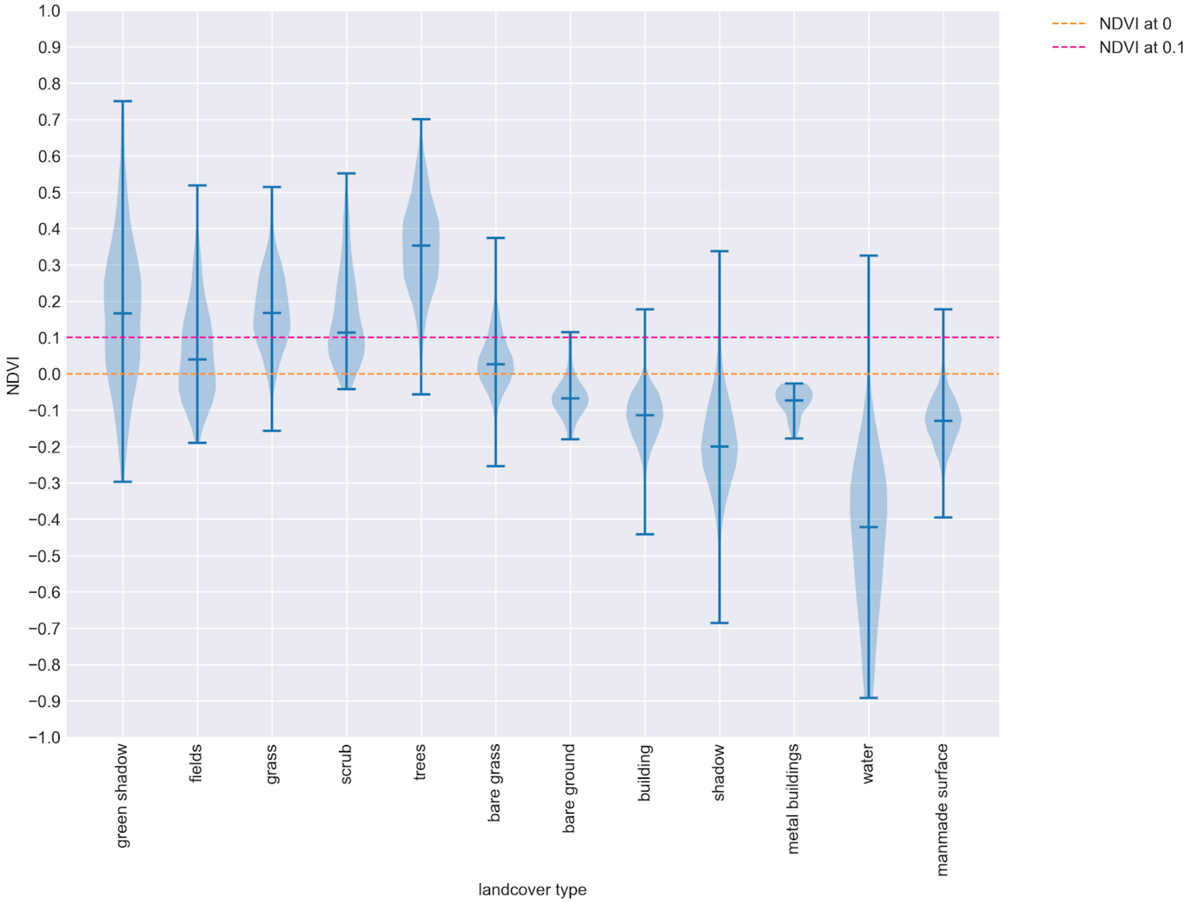
3. NDVI threshold and random sampling

Being concerned with extracting only green cover pixels from the aerial imagery, a 'threshold' NDVI value was needed.

Approximately 2,000 random sample points were generated within the GLA boundary. Each point was compared visually with the underlying aerial imagery and then manually assigned a land cover type from a set list of predefined classes e.g. fields, trees, building. Once 2,000

points were assigned a land cover type, a 10% random extract of these were checked again by a different user to ensure the land cover labels were assigned correctly.
These points were then intersected with the NDVI layer, extracting an NDVI value for each random point. These are visualised in the Chart 1 below.

Chart 1: NDVI values for various land cover types



Source: GLA GIS Team

Vegetation land cover types generally have a higher NDVI value than non-vegetation land cover types, which is as expected. Typically, trees exhibit the highest 'greenness' scores, with other vegetation types such as fields, grass, bare grass and scrub having lower values. The reasons for this are discussed in *Limitations.* The point at which a pixel's NDVI value was most likely to be vegetation ranged from between 0.05 and 0.1. Therefore, it was decided to create two threshold outputs at 0.05 and 0.1 and compare the findings.

For each tile, and for each threshold, Python GDAL and NumPy packages were used to isolate only pixels containing NDVI values equal or greater than the threshold. For simplicity, and to improve performance, pixels above the thresholds were assigned a new value of 1 and pixels below the threshold a value of 0. The resulting raster arrays were exported as a GeoTIFFs.

4. Converting NDVI raster tiles to vector

The 'NDVI threshold' GeoTIFFs needed to be converted to a vector format for use with additional land cover data, as well as for calculating areas and further analysis.

Python GDAL library was used to 'vectorise' each tile, by applying the Polygonize function. This was by far the most time-consuming part of the project, and also a trade off with using the GeoTIFF format rather than a more compressed output image format.

Each output vector was converted into an ESRI Shapefile format, with only areas of the tile exceeding the threshold retained.

Map 5: Example NDVI vector output for the 0.05 threshold

5. Preparation of additional green and blue land cover data

Custom queries were applied to OS MasterMap and Verisk GeoInformation UKMap in order to extract green and blue areas into separate datasets. Each extract was kept separate according to data source, as it was likely that there would be overlaps in some areas. The advantage of using two landcover data sources was the increased likelihood of picking up more green and blue areas that might only have been present on one of the datasets. A blue cover layer was extracted from the OS MasterMap layer, whilst separate green and blue cover layers were created from the UKMap dataset.

Each green and blue cover extract was tiled and labelled in line with the 2km by 2km grid used for aerial imagery. To avoid any overlap and remove duplication, the OS MasterMap blue cover extract was used to clip the UKMap blue cover extract on a tile by tile basis. These two blue cover extracts (MasterMap and UKMap) were merged into the same dataset to create a complete blue cover layer for London.

This blue layer was then used to clip the green cover layer, again tile by tile. Blue cover was given a higher precedence to green cover, therefore overlapping green cover areas would be clipped in favour of blue cover e.g. tree canopy overhanging a water body was removed and classified as blue. Reedbeds and other aquatic vegetation were also classified as blue.

6. Combining NDVI tiles with additional green and blue cover data

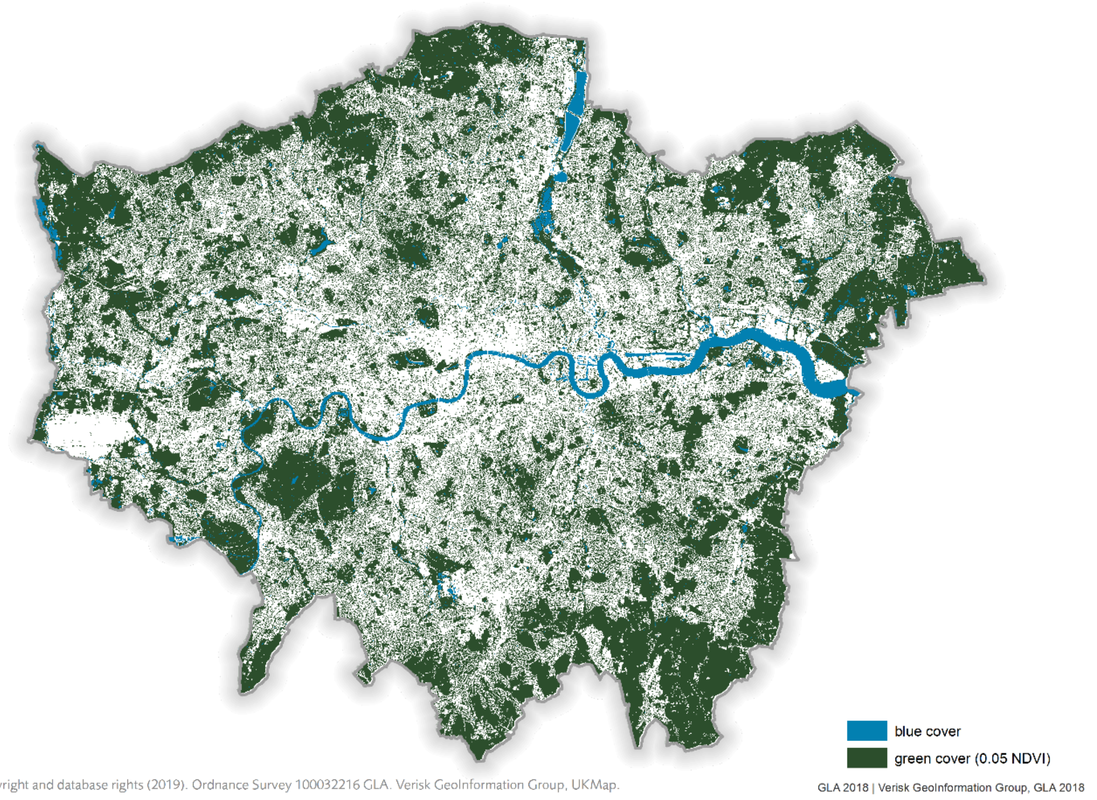
NDVI vector tiles were clipped by the additional land cover vector tiles (green and blue), to ensure no overlapping of polygons, which would otherwise cause inaccuracies with area calculations. The clipped NDVI vectors were then merged back with the land cover tiles to produce a comprehensive green and blue cover vector tile, comprised of NDVI analysis as well as known green and blue land cover.

Since known open spaces such as parks and playing fields often produced low NDVI values (see *Limitations* for more details), it was decided to allow the additional landcover polygons to take precedence. Therefore, if a green land cover polygon intersected with a NDVI polygon, the land cover layer was kept.

The NDVI polygons did, however, capture tree canopy, grass verges, green roofs and other smaller areas of vegetation that a commercial land cover dataset was unlikely to include. So, combining these two green layers would provide a better estimate of true green cover.

This processed was conducted for both NDVI threshold (0.05 and 0.1) tiles, so two main outputs were produced (Map 6).

Map 6: Merged NDVI with green and blue land cover layers – the NDVI 0.05 layer

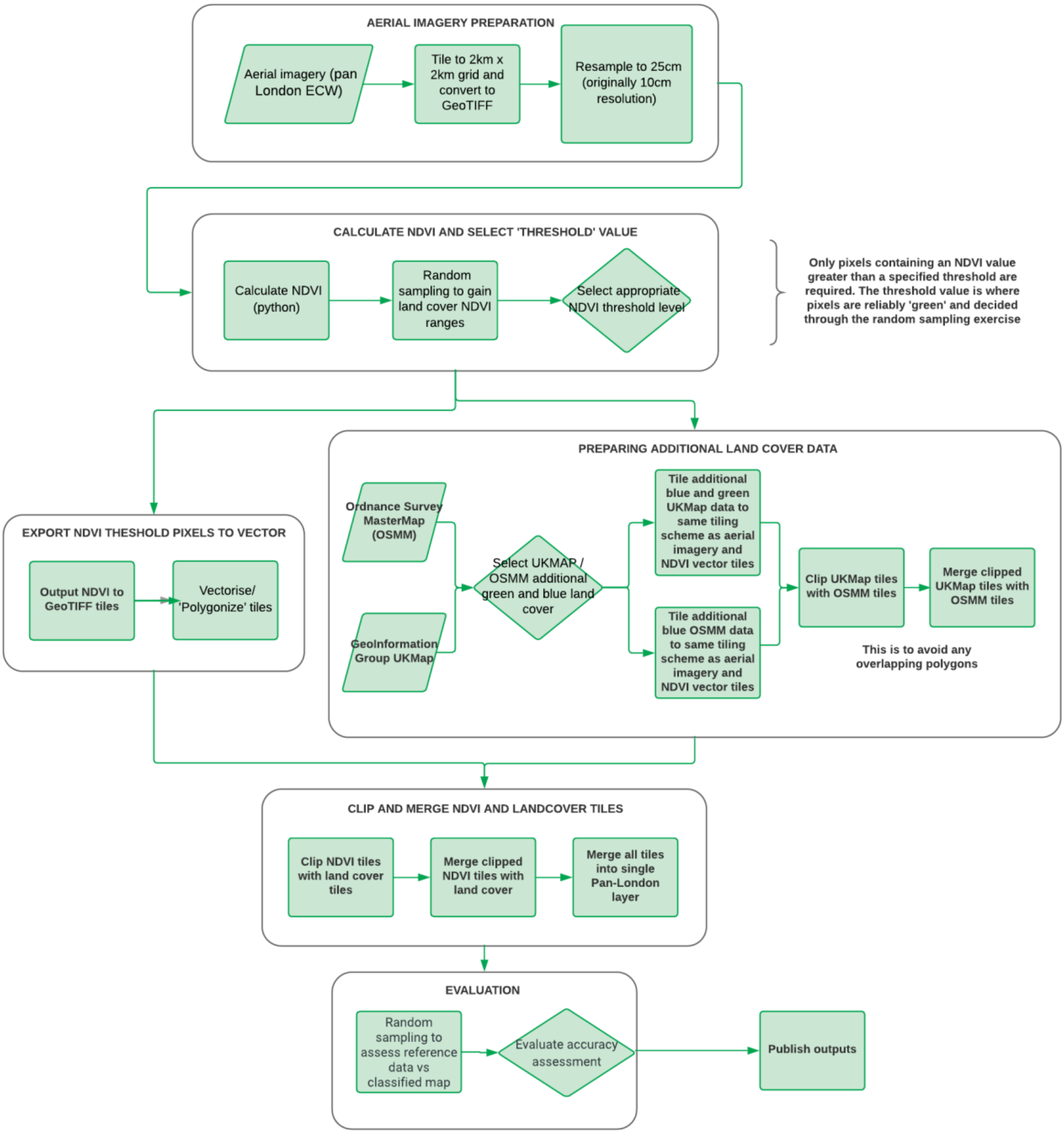


7. Calculating area

Each data source (NDVI, green, blue) making up each vector tile could be distinguished within its related attribute table. This allowed for area calculations to be applied at the land cover level, so an area for both green and blue cover could be calculated.

Areas were calculated for each of the 467 output vector tiles, and percentage land covers derived. These are presented in the *Results* section of this document. All areas are calculated using the British National Grid projection.

Figure 2: Detailed green/blue land cover map methodology



Source: GLA GIS Team

8. Accuracy assessment

Following any image classification project, it is important to carry out a quantitative assessment of the results. This not only provides the user with some certainty into the quality of the analysis but helps comparisons with other similar projects. An accuracy assessment compares the classified image, e.g. areas that are labelled as green or blue as part of this process, with another data source or 'ground truth' data.

This ground truth data came in the form of a manual, visual check of the aerial imagery by people. This comparison data was created by generating approximately 4,000 random points across the Greater London area. For each point, a record of the land cover class indicated by the underlying reference map (i.e. the aerial photography) was captured. This classification process is then repeated for all points. In this case we were simply interested in whether a point was correctly classified as green or blue, so recorded "green", "blue", "other" (i.e. not green or blue) or "unknown" (i.e. cannot clearly identify the landcover) from the imagery. The result is a table with two columns, a sample of which is shown in Table 4.

Table 4: Sample of ground truth point classifications of random samples

point_no	Reference data (aerial imagery)	Classified (new green and blue layer)
1	green	green
2	other	other
3	unknown	unknown
4	other	other
5	green	other
6	green	green
7	blue	blue

a point is classed as green in the imagery and falls within the green land cover layer

a point is classed as green in the imagery but falls outside the green land cover layer (e.g. classified as 'other')

This classified random sample was evaluated by taking a 10 percent sample of the points and checking the classification was consistent, any points found to be incorrectly classified were subsequently removed.

From this validation data it is possible to generate a confusion matrix that allows accuracy metrics to be calculated (Table5).

Table 5: Confusion matrix for the 0.05 and 0.1 NDVI with landcover data

NDVI threshold = 0.05	Reference Data		
	Green or blue	Not green or blue	**Total**

NDVI threshold = 0.1	Reference Data		
	Green or blue	Not green or blue	**Total**

Classified Data	Green or blue	2,112	131	2,243
	Not green or blue	25	1742	1,767
	Total	2,137	1,873	4,010

Classified Data	Green or blue	2,038	221	2,259
	Not green or blue	25	1,726	1,751
	Total	2,063	1,947	4,010

Source: GLA GIS Team

Producer's Accuracy is the map accuracy from the point of view of the map maker (the producer). This represents how often a real feature on the ground is classified correctly. In this case, it is the probability that an area of green, blue or other (i.e. not green/blue) in London has been correctly identified in the final land cover layer compared to the aerial imagery. This can be thought of as how well the classified data captures as much green or blue cover as possible.

The Consumer Accuracy is the map accuracy from the point of view of the map user, not the map maker. This represents how often the classification on the map will be present on the aerial imagery i.e. it does not incorrectly classify a point. In this case, if the map predicts a point to be green or blue, Consumer's Accuracy is the probability of that point being correct classified. This can be thought of as how well the classified data identifies green or blue cover without incorrectly overestimating it.

(Source: http://gsp.humboldt.edu/olm_2015/Courses/GSP_216_Online/lesson6-2/metrics.html)

The Producer's and Consumer's Accuracy results are shown in Table 6.

Table 6: The accuracy of the final green and blue cover layers

Output	Producer's accuracy %	Consumer's accuracy %	Overall accuracy %
NDVI 0.05 threshold + landcover	98.8	94.2	96.5
NDVI 0.1 threshold + landcover	98.8	90.2	94.5

Source: GLA GIS Team

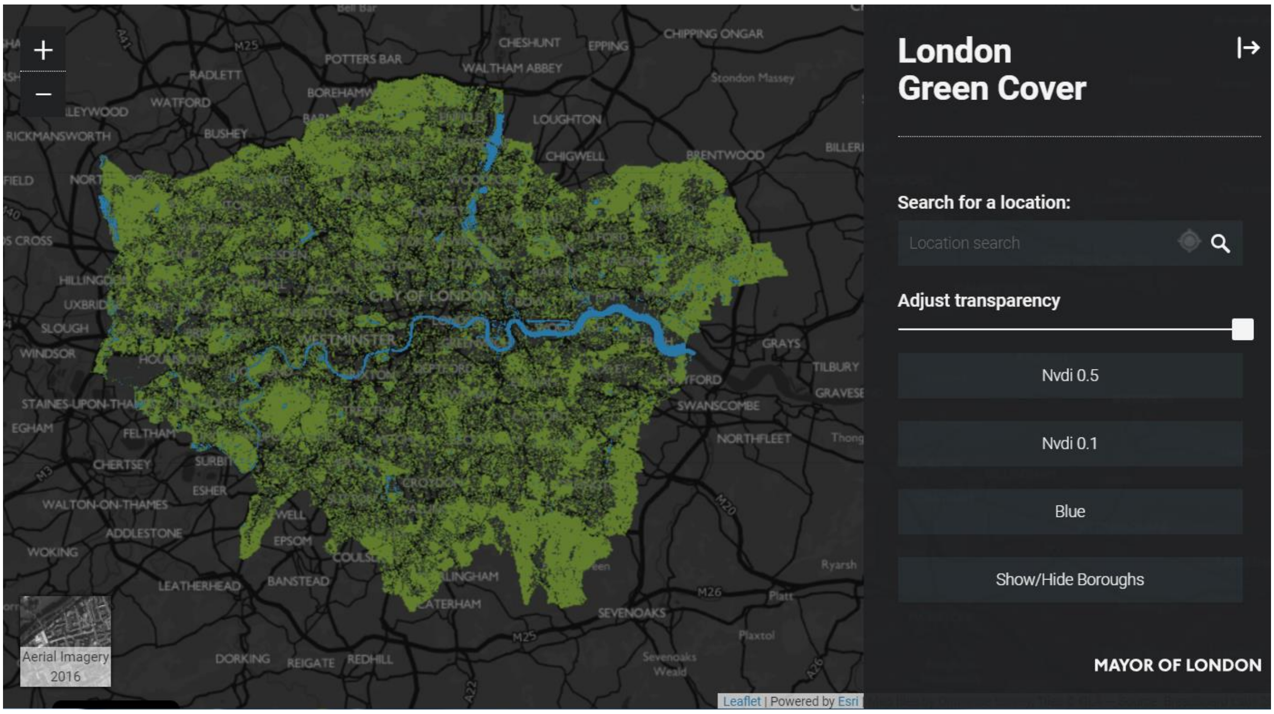
In both NDVI threshold scenarios the Producers Accuracy hold up very well, with scores above 98%. As expected, the Consumer's Accuracy is higher for the 0.05 NDVI threshold than the 0.1 value. This is attributed to unhealthy vegetation following a long, hot-summer, which would generate a weak NIR reflectance and a lower NDVI score. Lowering the NDVI threshold allowed many of these green areas to be captured at the 0.05 value further than being removed at the 0.1 threshold (Map 7).

Map 7: NDVI 0.05 versus 0.1 thresholds classifications, the lower threshold classifies more pixels as "green"

Results

The output from the above methodology resulted in high-resolution vector layers (one for each threshold) that estimated the green and blue land cover for Greater London. These layers (green NDVI 0.1, green NDVI 0.05 and blue) are visualised individually in a web map (Map 8).

Map 8: Green (and blue) cover web map



Source: https://maps.london.gov.uk/green-cover

The map allows users to view and interact with the green and blue layers, along with the source aerial imagery.

These layers allow percentage cover figures to be calculated for London as a whole, as well as other areas of interest (e.g. administrative wards and London Boroughs). Table 7 shows a summary of the results at the London level.

Table 7: Percentage green and blue land cover for London (with accuracies)

Output	Area of green and blue cover (ha)	% blue + green	% green	% blue (water)	Producer's accuracy %	Consumer's accuracy %	Overall accuracy %
NDVI 0.05 threshold + landcover	85,434	53.6	50.7	2.9	98.8	94.2	96.5
NDVI 0.1 threshold + landcover	80,506	50.5	47.6	2.9	98.8	90.2	94.5

Source: GLA GIS Team

All calculations were carried out using British National Grid projection (metres).

The variable NDVI threshold approach means the final green and blue land cover value can be estimated as a range between 50.5 and 53.6 percent. Removing the blue cover portion, the green cover value can be presented as a range between 47.6 and 50.7 percent. Although both overall accuracies are good, the 0.05 threshold represents a more reliable estimate, given the higher 96.5 percent score.

Although not available at the time of writing, it is the GLA's intention for summary statistics to be created for other geographies (Lower Layer Super Output Areas, Wards and London Boroughs. These outputs will be published on the London Datastore in due course.

The full detailed vector output layers created in this analysis are available from the London Datastore and published as open data for reuse.

Limitations and Issues

This is the GLA's first attempt at producing a detailed green cover layer for London. Although relatively simplistic in approach, the methodology builds on previous studies and is the best estimate currently available. However, there are clearly limitations to the methodology, these are discussed below.

Bare ground and eroded vegetation

NDVI analysis works well in identifying areas of healthy vegetation. Due to the time of year of aerial imagery capture (September 2016), vegetation was somewhat dry, and in many high use areas such as public parks, were partially eroded and large areas of bare grass can be seen. This would clearly have affected the spectral signature of these areas, causing them to be in correctly classified and missed by at thresholds set.

Map 9: Sample area of unhealthy vegetation and the NDVI 0.05 green cover classification layer

Ploughed agricultural areas

Equally, agricultural areas such as fields were sometimes ploughed (Map 10), and therefore devoid of any vegetation. Again, this would produce insufficient spectral signal to generate a positive NDVI index. Since the purpose of this analysis was to produce a practical map of green cover, it was decided that areas that are *usually* green would be most beneficial to capture.

Map 10: Sample area of ploughed fields being poorly identified by the NDVI index

Shadow
Shadow effects are noticeable in the imagery, often exacerbated by a low sun, creating large shadows later in the day. In these areas it is not possible to generate a positive NDVI score, so green cover capture is not possible though imagery analysis alone.

Unfortunately, any imagery will exhibit similar characteristics to those described above. To minimise these impacts, areas of green and blue cover were also selected from alternative survey-derived land cover datasets (Ordnance Survey MasterMap and Verisk GeoInformation Group UKMap).

Blue cover
Blue cover layers were extracted from Ordnance Survey MasterMap and Verisk GeoInformation Group UKMap datasets. In both sources, the layers extracted were similar in coverage but not identical, some blue areas were not recorded in one dataset but were present in the other. So, it was decided that a combination of both would provide the most accurate estimate.

Green cover
A proportion of this layer was generated from the UKMap Feature Type Code (FTC). In many cases it accurately filled in gaps missed by the classified imagery (e.g. bare grass/shadow) and was chosen as the definitive green layer. However, in some scenarios the FTC classification proved unreliable e.g. school playing fields (Map 11), and although a compete polygon was classified as ‘green’, there were clearly non-vegetated areas, such as outbuildings and paved

areas, within it's boundary. In these cases, a decision was made to exclude these FTC codes from the selection and avoid over estimating green cover.

Map 11: Sample secondary school area classified as 'vegetated' but containing paved areas and a building.

There were also some border-line cases with certain FTC codes where an area may not be 100% green cover, such as allotments and cemeteries, but were included in the green selection (Map 12). These codes and descriptions are provided in Appendix I, so it's clear what has been included.

Map 12: Example of a cemetery classified as green land cover from UKMap

Temporal change

All data sets represent snap shots at different points in time (2016), clearly in some areas, land cover types will have changed since the time of survey and not be reflected. All data sources are currently maintained and updated by respective data providers, so an updated future study would be possible (see *Future Work*).

Conclusion

The method presented here represents a simple but robust approach to estimating London's green and blue cover. The overall maps accuracies in Table 7 indicates that a map user may be confident in the green and blue cover layer created but aware of the issues such as shadow, unhealthy vegetation and temporal change may introduce.

This work compliments previous studies and, for the first time, provides a detailed high-resolution green cover layer for Greater London that can be used for further analysis.

Future Work

Publishing the output data, methodology and processing scripts allows for similar studies to be replicated in future. The GLA intend to repeat this study every five years and begin to investigate temporal change in London's green and blue cover.

The approach to this analysis was relatively simplistic and the quality of imagery data would allow for more sophisticated Machine learning-based approaches to be applied. This could undertake further image processing techniques that generate additional image features (e.g. additional vegetation indexes, neighbourhood analysis and image texture) and not simply the NDVI in isolation. These features can then be used collectively to predict pixel-level vegetation information using a machine learning classification algorithm.

Combining the UKMap data with other sources, such as Ordnance Survey MasterMap (Topography and Greenspacelayer), could improve some of the inaccuracies in the green cover classification. By removing non-green Mastermap polygons from the UKMap green cover layer, more of this dataset could be used to improve areas where the NDVI analysis was problematic (Map 13). Furthermore, integrating with data sources such as the Curio Tree Canopy layer and LiDAR (Light Detection and Ranging) datasets could further refine the quality of the output layer.

Map 13: OS MasterMap Buildings layer identifying an incorrectly classified vegetated area in UKMap

During this study we consulted with organisations with expertise in this field of work such as; Greenspace Information for Greater London CIC (GIGL), Verisk Analytics GeoInformation Group and Breadboard Labs, to help establish a sound and robust approach to the project. We are keen to hear from other interested parties on how the methodology could be improved and refined further. We're keen to hear of any additional analysis/project undertaken using the green cover vector layer available to download from the London Datastore.

Appendix I

A list of Verisk Geoinformation UKMap land cover selections used to create the green cover layer:

Feature Classification Codes (FCC) for green cover:

FCC1	FCC2	FCC3	FCC4	Description
11	6			Cultivated places, Arable farm places, Horticultural places,Allotment gardens, Glass house,Orchard with grass
11	8			Non cultivated places,Grazing places, Permanent pasture, Rough grazing, Forestry places, Coniferous forest, Coppice,Deciduous forest, Mixed forest, Tree nursery, Field margin - hedgerow, Field margin - line of trees, Field margin - undifferentiated, Field margin - grassland,Field margin - hedgerow with trees, scrub land
14	2	2		Nature reserve, Site of special scientific interest
15	1	1		Zoological garden, Country park, Gardens (not private), Park, Picnic site, Recreational open space, Botanical garden
15	3	1		Ball game pitches and grounds, Association football ground, Cricket ground, Rugby football ground
15	3	3	2	golf course
24	7	2	1	cemetery

A secondary selection was also made by Feature Type Code (FTC) code to select "Vegetated" feature types from these FCC codes.

Feature Classification Codes (FCC) for blue cover:

FCC1	FCC2	FCC3	FCC4	Description
15	4	2	1	Boating facilities
15	4	3	1	Rod/recreational fishing place
22	3	1	1	Canal
22	3	1	2,3,4	River,Marina,Mooring
23	1	3	4,5	Pond or lake, water course
23	3	1	1	Reservoir

Additionally, a selection was also made by Feature Type Code (FTC) code to select "Water" features types.

GLAINTELLIGENCE

Greater London Authority
City Hall
The Queens Walk
London SE1 2AA

Tel 020 7983 4000
Fax 020 7983 4674
Minicom 020 7983 4458
Email intelligence@london.gov.uk

https://www.london.gov.uk/what-we-do/research-and-analysis

MAYOR OF LONDON