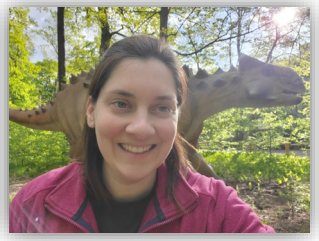
Generating and sharing a two-metre resolution land cover map of London

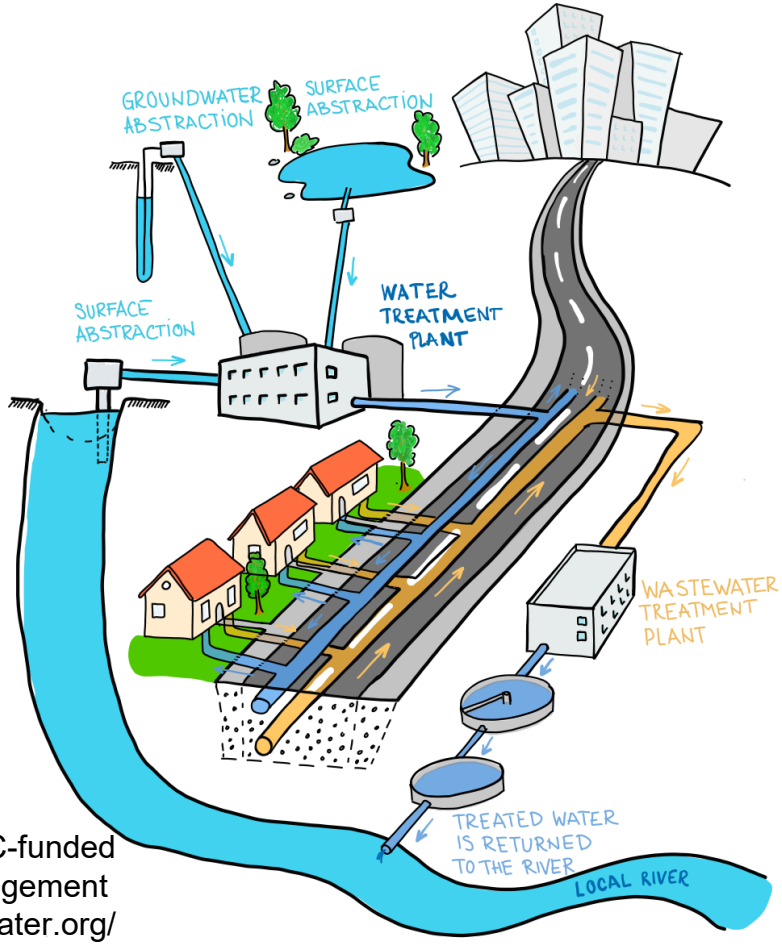
DR CHRIS JACKSON

British Geological Survey



Tijana Jovanovic, PhD

- Urban hydrologist
  - with background in civil engineering
- Research interests
  - urban heterogeneity, complexity, and sustainability
  - green infrastructure
  - Socio-hydrological resilience

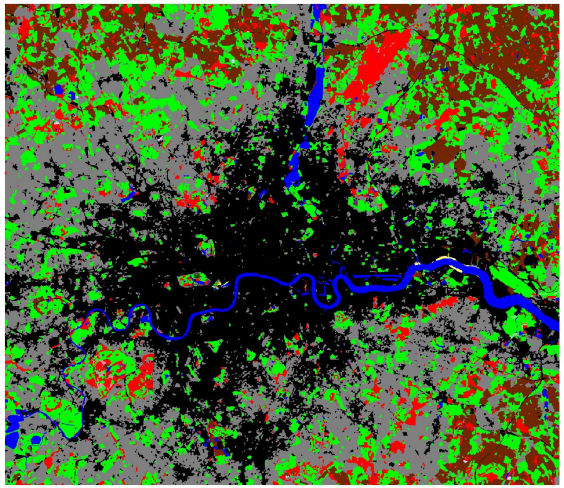


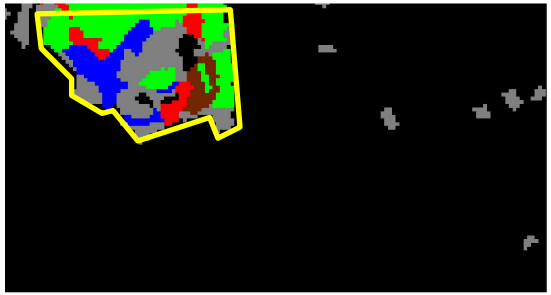
CAMELLIA
Community Water Management for a Liveable London

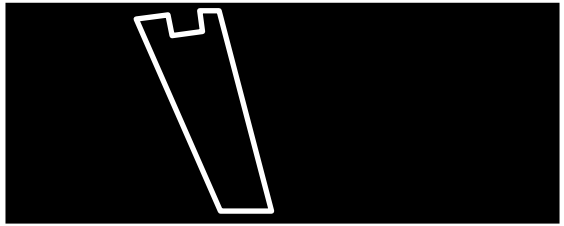
This work was developed under UKRI NERC-funded CAMELLIA project (Community Water Management for Liveable London) https://www.camelliawater.org/

BGS

Existing
freely available
land cover datasets
are not
of sufficient resolution
for
urban studies







CEH 25m Landcover

reshaped_LC2m

- 0
- 1
- Coniferous Woodland
- 3
- Improved Grassland
- Neutral Grassland
- 6
- 7
- Fen, Marsh and Swamp
- 9
- 10
- 12
- Saltwater
- Freshwater
- 16
- Littoral rock
- Littoral sediment
- Saltmarsh
- Urban
- Suburban

Google Satellite

Existing freely available **land cover** datasets **are not** of sufficient resolution for urban studies

Blue Green Infrastructure is designed to be fitted in gardens and patches along streets. To be able to model them one needs detailed information of the land cover

What datasets do we have? More importantly can the derived datasets be published?

What datasets do we have? More important can the derived datasets be published?

OS MasterMap Topography Layer

Technical information

- Very detailed topography layer, but without details about vegetative land cover
- Proprietary dataset, but with some layers released as open access (buildings, roads...)
- Can be used under the Public Sector Geospatial Agreement

What datasets do we have? More important can the derived datasets be published?

- Impervious areas such as buildings, structures, roads are covered, but for vegetation a near infrared sensor would be needed
- PGA near infrared areal imagery at resolution of 50cm, to calculate Normalized Difference Vegetation Index (NDVI). NDVI helps with identification of green areas
- Green colour (and NDVI) represents both trees and grass, to distinguish between the two, we
  - Digital Surface Model (DSM)
  - Digital Terrain Model (DTM)
  - Height = DSM – DTM
  - Various DSM and DTM products available (from different suppliers) at different resolutions (1m to 5m) and legal attachments

All the datasets are there… Which software to use?

- Many known GUI software (ArcMap, ArcGISPro, QGIS, SAGA GIS, Whitebox Geospatial…) crash and/or start lagging, just by opening large datasets and can take ages to process them

- Code to process the landcover was developed in Python (within Jupyter Notebooks) using native packages (main were Rasterio and Geopandas, other included os, Numpy and Shapely, and for visualisation Matplotlib and Seaborn)
- Some of the operations used in the workflow:
  - Reshaping to raster resolutions of 2 m
  - Cutting raster and vector datasets to 1km OSGB tiles
  - Extracting datasets stored in an ESRI geodatabase
  - Rasterizing vector datasets
  - Extracting lists of available files from folders

All the datasets are there… How to do the processing?

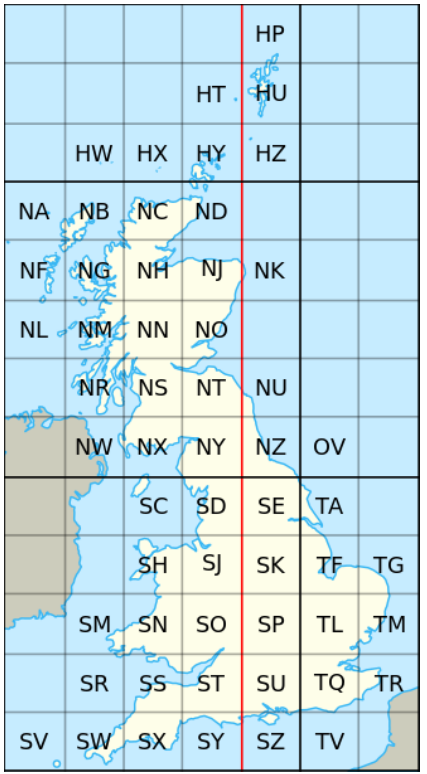
- Processing was done in tiles of 1 km OSGB (Ordinance Survey National Grid reference system). All the raster data was pre-processed (either reshaped and/or cut to this grid) and stored locally (except arial imagery which was already stored in 1km OSGB titles)
- OS master map is a vector dataset, it was cut into same extent as raster data and rasterized
- Within Greater London Area (GLA) and Wastewater Resources Zones of London there were total of 2,349 1km OSGB tiles
- Processing time for the whole dataset is about 2 hours.

NOTE: calling for standardising the way data is stored on company servers

- The project would have been completed a lot quicker if all the DSM and DTM data sets were stored in the same OSGB tile.

- Licencing and data quality concerns meant that nearly every available dataset to us (DTM, DSM combination) had to be processed differently, tiles the datasets were stored in:
  - 100km OSGB tiles
  - 50km OSGB tiles
  - 10km OSGB tiles
  - contiguous for the whole UK

[https://en.wikipedia.org/wiki/Ordnance_Survey_National_Grid]



NOTE: Jupyter Notebooks as your digital “lab” notebook

File Edit Selection View Go Run Terminal Help

_LC_data_OSMasterMapTopoAreas.ipynb × _LC_data_18grid_workflow_50cm.ipynb

CAMELLIA_HydroModels > Projects > london_LandCover_TJ > Code > _LC_data_OSMasterMapTopoAreas.ipynb > Creating a test file for the

+ Code + Markdown | Run All Clear All Outputs | Outline ··· Select Kernel

Creating a test file for the topographic areas

Testing if the extraction worked and I have a fully functioning dataset

Problem When cropped the geodataframe converts some polygons to points and some to multypolygons This needs to be fixed in order to export the data. For now rather than fixing it I will extract only the polygons and use them to create the rest of the code

Solution : when cropping is finished take the geometry from the original file, for those that ended up as points or multipolygons

```
[1]
1  import geopandas as gpd
2  import pandas as pd
3  import numpy as np
4  import rasterio as rio
5  import os
6  import matplotlib.pyplot as plt
7  from mpl_toolkits.axes_grid1 import make_a
8  from rasterio.merge import merge
9  from rasterio.plot import show
10 from rasterio.mask import mask
11 from rasterio.plot import plotting_extent
12 from rasterio.enums import Resampling
13 from shapely.geometry import box
14 from rasterstats import zonal_stats
15 from earthpy import clip as cl
16 import earthpy.spatial as es
17 import earthpy.plot as ep
```

Decision

For now will go with the Make and Descriptiv classification. It seem that the "Make" column helps to easier distinguish between natural and manmade structures which is what we need. Where we would need help is in the cases of "Make" = " Multiple" where these can be either natural or artificial surfaces, this part can be distinguished from ndvi, imagery, DSM, and DTM

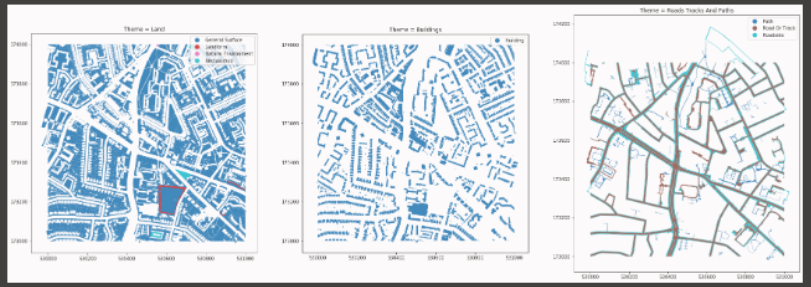
File Edit Selection View Go Run Terminal Help

_LC_data_OSMasterMapTopoAreas.ipynb × _LC_data_18grid_workflow_50cm.ipynb

CAMELLIA_HydroModels > Projects > london_LandCover_TJ > Code > _LC_data_OSMasterMapTopoAreas.ipynb > Creating a test file for th

+ Code + Markdown | Run All Clear All Outputs | Outline ··· Select Kernel

[32] Python



```
[35]
1 #exporting the image
2 fig2bPath = f'{pathOutput}/fig2b_exploringTopo_Theme_Descriptive_columns.png'
3 fig2b.savefig(fig2bPath,bbox_inches='tight')
```

Python

on. It seem that the "Make" manmade structures which is what Make" = " Multiple" where these e distinguished from ndvi,

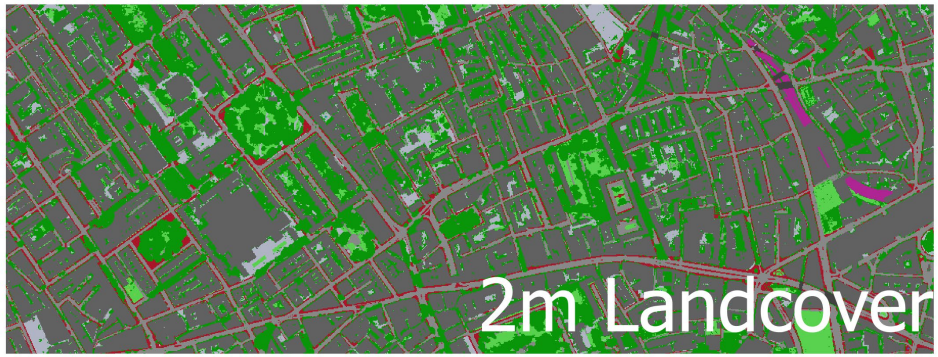
Notes on publishing and some dataset caveats

- We have been in conversation with many legal teams (different providers) and emailing back and forth took a lot of our time and resources.

- We are grateful to the EA for releasing their LIDAR derived DSM and DTM. Without it, publishing this dataset wouldn't be possible (other providers were not keen on allowing us to publish under open government licence).

- We are also grateful to Ordinance Survey, who were open to dialogue and after some landcover class ‘tweaking’ will allow us to publish under open government licence.

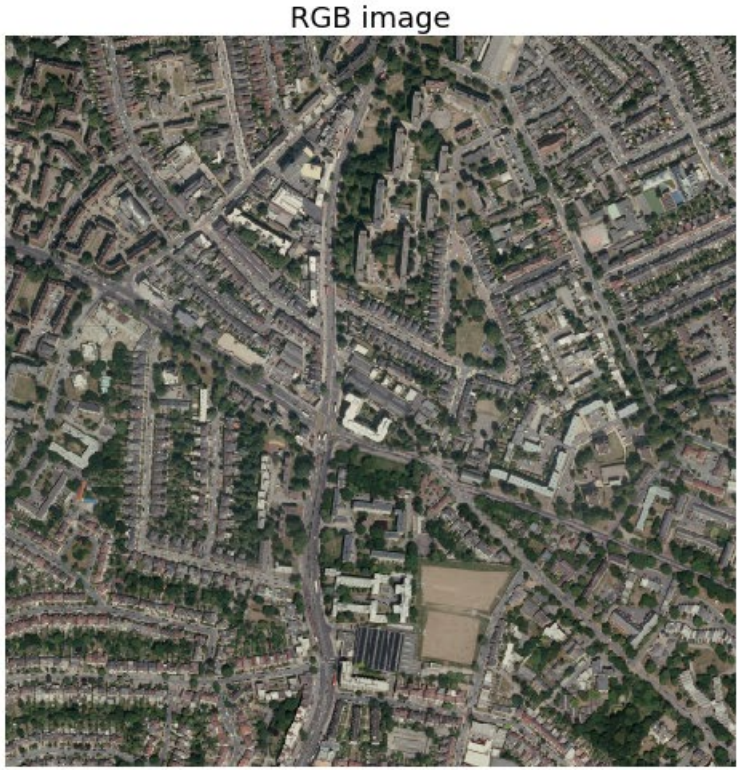
- To create this dataset, we worked with datasets at our disposal. This means that while buildings and roads are created from up-to-date data, the vegetation and street trees are created from images from a range of years (2008-2010).

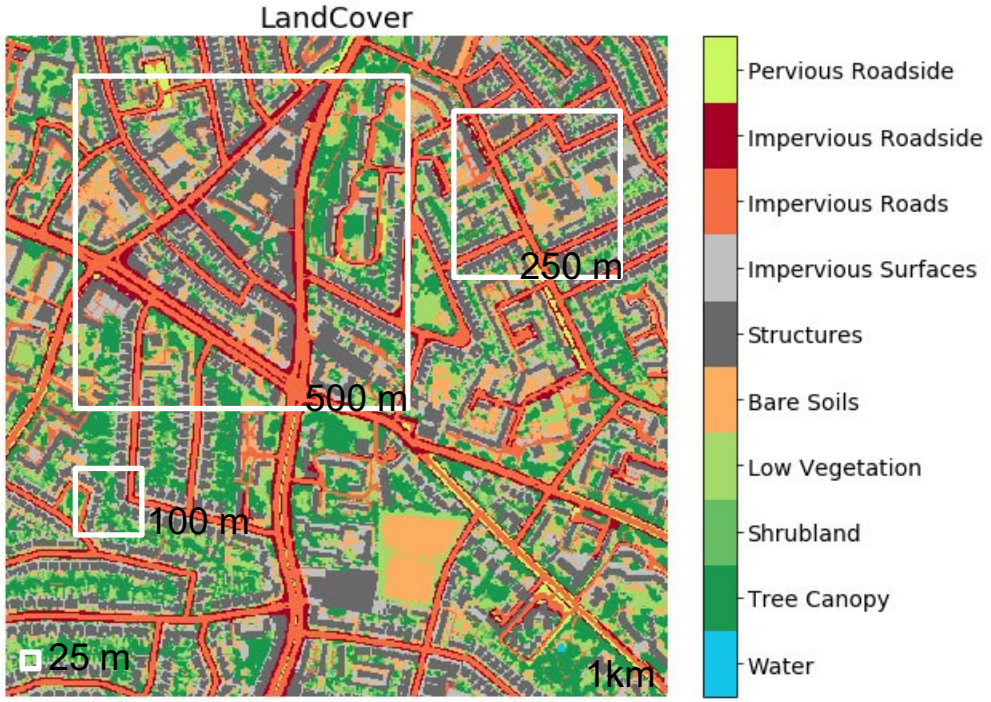


2m London Landcover

- No data
- Water
- Tree canopy
- Low vegetation
- Bare Soils
- Buildings
- Impervious Surfaces
- Impervious Roads
- Impervious Roadside
- Pervious Roadside
- Tree over Impvervious Surfaces
- Tree over Impervious Road/Roadside
- Tree over Pervious Roadside
- Rail
- Bridge/Structure
- Tidal Water

The final product





Thank you!

Contact email: tijj@bgs.ac.uk

The final product