A data-driven approach to creating a Built Environment Context and Change Atlas (BECCA)

Jessica Hepburn[*1], Rich Mitchell[†1], Jenna Panter[^2] and Fiona Caryl[‡1]

[1] MRC/CSO Social and Public Health Sciences Unit, University of Glasgow
[2] MRC Epidemiology Unit, University of Cambridge

January 16, 2023

Summary

Built Environment Context and Change Atlas (BECCA) aims to produce an online map quantifying built environment change between 2016 and 2020 across Great Britain. Composite variables will be created using a range of data sources to describe what type of place an area is and how this has changed through time. Linking this data to individuals' health in turn can highlight influences as well as direct future planning and policy. BECCA will available online on a public-facing website for researchers, the public, and private organisations to use to help support decision-making.

KEYWORDS:

Environmental Change, Public Health, Multivariate Analyses, Decision Making

1. Introduction

The Built Environment Context and Change Atlas (BECCA) is a collaboration between the University of Glasgow, Imperial College London, and the University of Cambridge, aiming to produce an online interactive atlas of environmental conditions and changes for the whole of Great Britain (GB). The atlas will be available online on a public-facing website for researchers, the public, and private organisations to use. This project has (and will continue to) collaborate with several stakeholders from private organisations, charities, and local and national governments. This collaboration is to ensure that BECCA is both user-friendly from a user experience perspective and that the content is relevant to decision and policy makers.

The UK has some of the world's best longitudinal data on the health and lives of individuals but lacks comparable consistent longitudinal environmental data (UK Data Service, 2022), and those that exist are particularly hard to access for non-experts. This matters because the environment is an important determinant of health (World Health Organisation, 2019). Researchers will be able to use BECCA to link changes in the built environment with changes in health or behaviours to evaluate the impacts of the environment. Local policymakers might want to identify areas subject to environmental changes, which could be policy implementation or investigation areas. A range of different variables will be considered, including, but not limited to, buildings, roads, greenspace, land cover type, air pollution and noise. Changes in these variables at a 4-year interval (from 2016 to 2020) will be calculated at a fine spatial scale (500m grid), as well as composite variables created to classify what type of place an area was at time point one, and how this has changed by time point two.

2. Methods

2.1. Data

2.1.1. Sampling units

As this project examines change, data sources and the associated variables had to link temporally to get a snapshot in time of mapped change. Each variable was summarised into a 500m grid which was based

[*] Jessica.Hepburn@Glasgow.ac.uk
[†] Rich.Mitchell@Glasgow.ac.uk
^ Jenna.Panter@mrc-epid.cam.ac.uk
[‡] Fiona.Caryl@Glasgow.ac.uk

on the Ordnance Survey (OS) gridlines in OSGB36 and was made up of approximately 940,000 cells. Each cell has an assortment of values, as summarised in Table 1 (below) across two-time points; 2015/16 and 2020. Input data needed to be comparable through time and generated through reproducible methods. The following sub-sections will discuss in detail the sources and choices for data within this project.

Table 1 Data implemented and their sources

Data Source	Product	Metric	Dates		URL
Ordnance Survey	OpenMap Local	Road length (m^2)	2016	2020	Access here
Ordnance Survey	MasterMap	Building Volume (m^3)	2015	2020	Access here
		Building Count			
Centre for Ecology & Hydrology	Landcover land parcels	Urban (m^2)	2015	2020	Access here
		Woodland (m^2)			
		Grassland and Arable (m^2)			

2.1.2. OpenMap Local

OpenMap Local (OML) is an open data product through the Open Government License from OS. Containing thematic layers, including buildings, roads, and woodland with new versions released every 6 months (Ordnance Survey, 2022). However, just because a new version is released this does not mean GB has been re-surveyed, with resurveying focused in urban areas. From examining the first available year of OML (2015) as well as the next consecutive year (2016) it was evident that there had been changes in the categorisation of data to include on the open product, this is demonstrated in Figure 1 with the road data layer. There is a substantial amount of mapped change between 2015 and 2016 compared to 2016 and 2017. The data became much more stable after 2016, and as one of the focuses of BECCA is change, it requires consistency in which features are mapped.

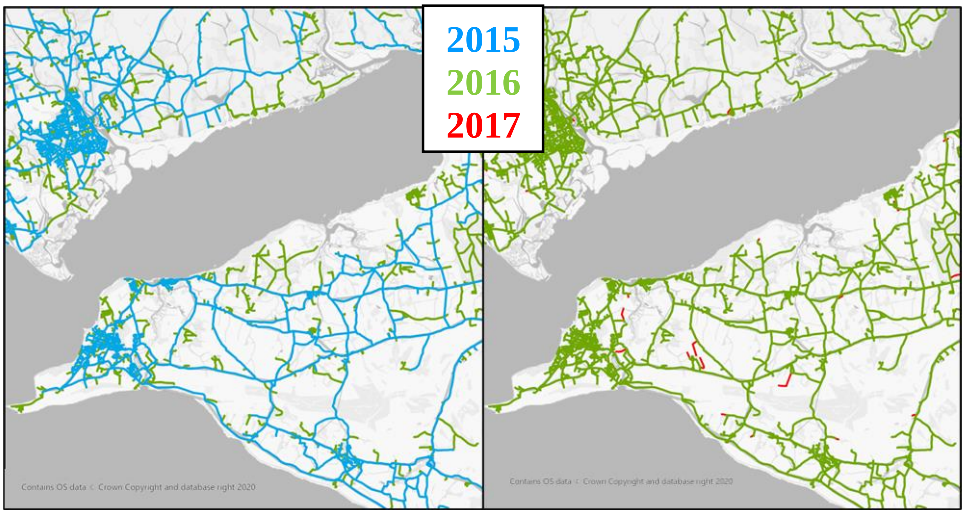

Figure 1 OpenMap Local Roads – 2015, 2016 and 2017 comparison

2.1.3. MasterMap

Alternately, MasterMap is an OS premium product, containing geometry for each building in GB (whereas OML can be an amalgamation of generalised building outlines). One of the main benefits of MasterMap is that each feature contains a Topographic Identifier (TOID) allowing for attributes to be linked. One beneficial joinable attribute is building heights; allowing for building volume to be calculated. This is a useful urban growth typology as can indicate whether development is upward (high-rise buildings) or outward (urban sprawl) (Boynton Tobey et al., 2019). The frequency of resurvey, particularly in urban areas, is high, as OS is the national mapping agency, allowing for mapped change to be captured relating to infrastructure (roads/buildings) from MasterMap data.

2.1.4. Landcover

Although OML does include some landcover information, such as woodland, when examining this further its updates did not seem to be that frequent or consistent, which is why OS data is not suitable for longitudinal assessment of natural land covers. Landcover data from the Centre for Ecology & Hydrology (CEH) quantifies landcover types across land parcels in GB. Broad classes from amalgamating detailed classes (such as coniferous and deciduous woodland to make woodland) were implemented based on the information produced by CEH. Their 5 aggregate *broad habitat* classes (*urban*, *woodland*, *water*, *grassland* and *arable*) are designed to be reproducible/comparable through time. In addition to the 5 classes, it was decided that *Grassland* and *Arable* would be combined in our analysis as these are likely to contain similar types of change (for example green field sites to new residential development). The landcover data is available in 2015 and 2020, although this is a different sequential start to the OpenMap Local input it is the most extensive and detailed landcover data available in GB. Landcover data was validated against the OS data to make sure that both data products captured known ground-truthed change across the two-time points, identifying any lags.

2.2. Preliminary analysis

To address the aim of this project, the metrics mentioned in the previous section need to be combined for each of the time points across the entire extent of GB. Combining metrics to classify cells will identify the type of place somewhere was, how it has potentially changed (such as urbanisation/deurbanisation), and what it has changed to. The main data-driven method implemented so far to classify cells is k-means; a multivariate unsupervised clustering technique requiring continuous variables as input. The k-means thus far have been conducted through ArcGIS Pro, outputted alongside the cluster id for each cell is the pseudo-f statistic which is the ratio of the between-cluster variation to the within-cluster variation (Milligan and Cooper, 1985), indicating the optimum number of clusters.

3. Results

3.1. K-mean cluster characteristics

When combing OS-derived data (*building volume*, *building count*, *road length*) alongside CEH landcover coverage (*urban*, *arable and grassland*, and *woodland*), 10 clusters are outputted which demonstrate a range of different types of place. Clusters A – E are predominantly urban with clusters F- J predominantly rural. Below, Table 2 shows the clusters identified and their characteristics, the colours used on this table will be reciprocated throughout this section.

Table 2 Clusters identified so far and their characteristics

Cluster ID	Characteristic
A	High-rise, largest buildings, central business district
B	High-rise, large buildings
C	Smaller buildings in high density urban (more likely to be residential)
D	Smaller buildings in moderate density urban with more green space
E	Rural area with infrastructure
F	Predominantly grassland/arable but with some infrastructure
G	Grassland/arable and woodland but with some infrastructure
H	Predominantly grassland/arable but with minimal infrastructure
I	Primarily woodland
J	Predominantly water

3.2. Cluster Classification across different geographies

The cluster classes can be used to understand the type of place a cell is. Figure 2 shows an expected distribution across GB geography, even at this scale able to make out major cities and towns. When we focus on Glasgow City and its centre these are all allocated to one of the urban classes, as expected, with the centre of town allocated to *Cluster A - High-rise, largest buildings, central business district*, with the image below demonstrating the urban morphology of Glasgow city centre.

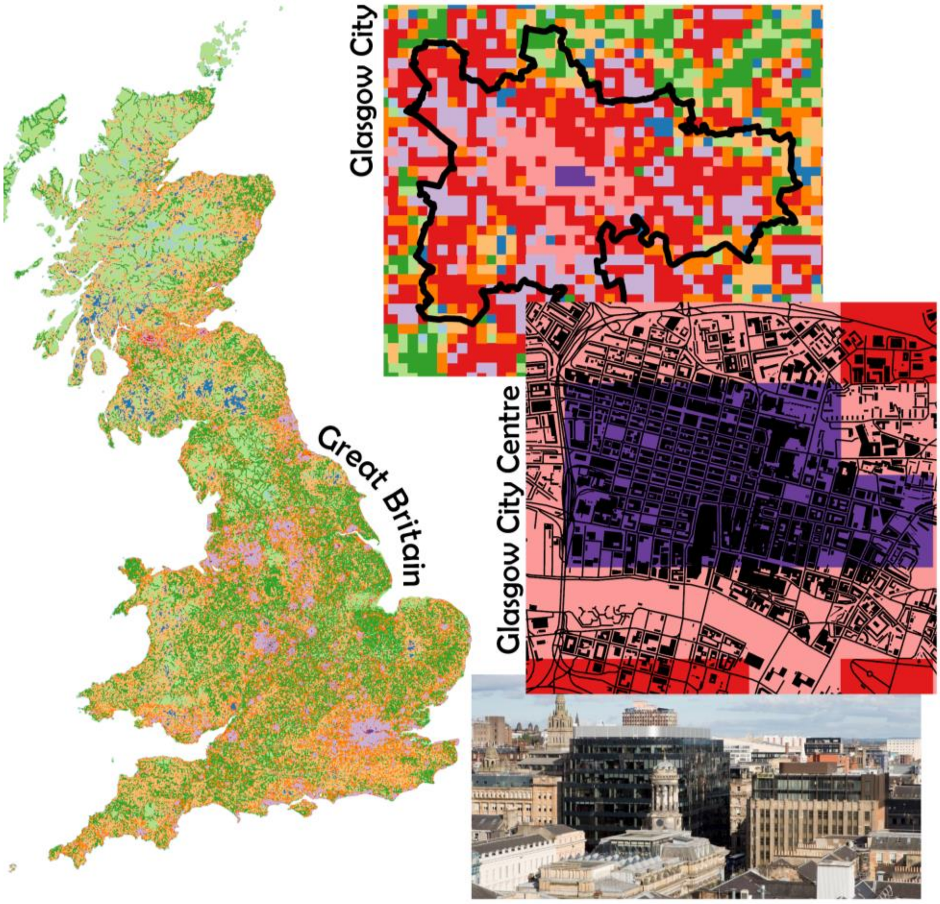

Figure 2 Maps showing clustering of metrics across Great Britain, Glasgow City and Glasgow City Centre for 2020

When comparing the percentage allocation of cells across Glasgow, Scotland, and GB for each cluster (Figure 3), it is evident that there are significant differences. Scotland has higher percentage of cells is *Cluster H - Predominantly grassland/arable but with minimal infrastructure*, *Cluster I - Primarily woodland and Cluster J - Predominantly water*, which would be expected due to the Highlands, Lowlands, and Lochs. In addition, 91% of Scotland's population lives in settlements and localities, which accounted for 2.3% of Scotland's total land area (National Record of Scotland, 2022), so this difference to overall GB would be anticipated.

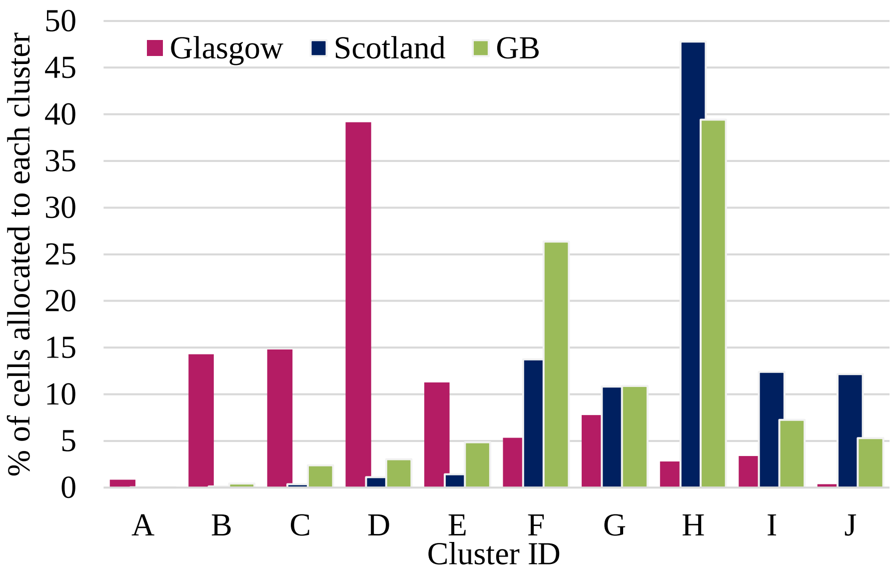

Figure 3 Percentage of cells allocated to each Cluster ID across Glasgow, Scotland, and GB

The interest is in whether these classifications of cells change with redevelopment/rejuvenation and new developments, which will be the next step in this project. So far preliminary results indicate that change can be detected from this clustering technique and a combination of input metrics.

4. Next steps

Although the analysis performed so far is promising, it identifies different types of places across the two-time points. Further work will examine these clusters in more detail as well as link them to health studies by exploring links between place, change and inequalities. The distribution of changes will be investigated to determine if certain types of place and types of change are more common in areas of more/less deprivation. Furthermore, detailed signature snapshots will be written including imagery (aerial/street view) of what each cluster typically represents. There will be ground-truthing against areas of known change to see if we are capturing it in a meaningful way. Additionally, to continued analysis, this project is closely linked to producing a public-facing website for stakeholders, the public and the government. Usability tests will be conducted on the product, making sure that it is both user-friendly and accessible.

5. Acknowledgements

This project is funded through the Medical Research Council national assets scheme.

References

Boynton Tobey, M. et al. (2019) 'Building up or spreading out? Typologies of urban growth across 478 cities of 1 million+', Environmental Research Letters, 14(12), p. 124077.
Available at: https://doi.org/10.1088/1748-9326/AB59BF.

Milligan, G.W. and Cooper, M.C. (1985) 'An examination of procedures for determining the number of clusters in a data set', *Psychometrika*, 50(2), pp. 159–179.
Available at: https://doi.org/10.1007/BF02294245/METRICS.

National Records of Scotland (2022) *'Population Estimates for Settlements and Localities in Scotland, Mid-2020'* Available at : /https://www.nrscotland.gov.uk/files//statistics/settlements-localities/set-loc-20/set-loc-2020-report.pdf

Ordnance Survey (2022) *OS OpenMap - Local*.
Available at: https://www.ordnancesurvey.co.uk/business-government/products/open-map-local
(Accessed: 16 January 2023).

UK Data Service (2022) *Trusted access and training to use the UK's largest collection of economic, population and social research data for teaching, learning and public benefit., UK Data Service*.
Available at: https://ukdataservice.ac.uk/
(Accessed: 16 January 2023).

World Health Organisation (2019) *Health, environment and climate change. WHO global strategy on health, environment and climate change: the transformation needed to improve lives and well-being sustainably through healthy environments*.
Available at: https://apps.who.int/iris/handle/10665/43457.

Biographies

Jessica Hepburn is a Research Associate at MRC/CSO Social and Public Health Sciences Unit within the Places and Health Programme at the University of Glasgow.

Rich Mitchell is a Professor at MRC/CSO Social and Public Health Sciences Unit within the Places and Health Programme at the University of Glasgow.

Jenna Panter is a Senior Research Associate at MRC Epidemiology Unit at the University of Cambridge School of Clinical Medicine.

Fiona Caryl is an MRC Skills Development Fellow at MRC/CSO Social and Public Health Sciences Unit within the Places and Health Programme at the University of Glasgow.