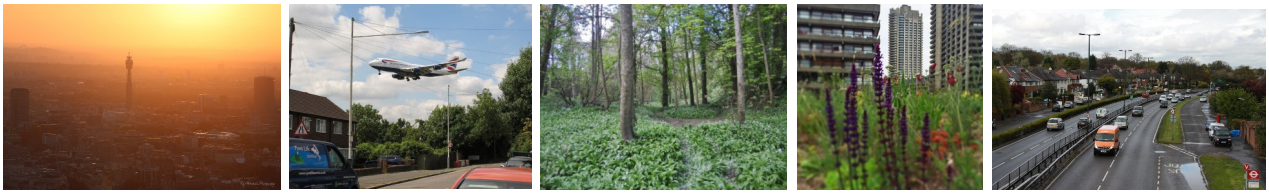
Green Poverty Index for England



This is a description of the method used to create a neighbourhood-level Green Poverty Index for England. It aims to capture environmental quality as it affects populations (both daytime and night time) across the country. The results are limited to England because of open data availability, but can be extended to include likewise datasets from other countries in the UK, and more recent data where it is available. The key themes included in the model are:

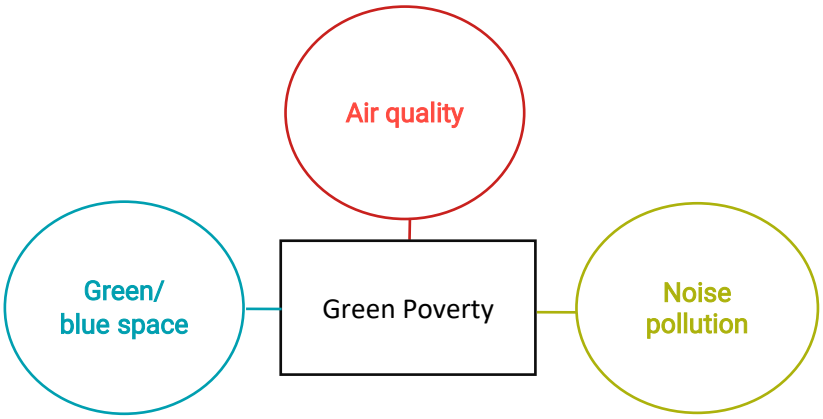


Figure 1: Key themes

Modelling overview

The model uses the 1km x 1km grid squares as its scale; that of the existing government modelled air pollution background pollutant data. Data is aggregated to these grid squares, with a score and rank of environmental quality. This may be considered 'neighbourhood level': the measures are not specifically residential, and so the environmental quality in any one grid square may have an effect on both 'daytime' and 'night time' populations, or in other words the resident neighbourhoods in which people live, but also those local places that people move through for multiple reasons (commuting, shopping, leisure, school and work and generally moving through an area for example).

In order to select appropriate data, metrics of each input dataset were considered for coherence against the themes, along with an appropriate available spatial scale. The ideal datasets for modelling do not always exist, therefore the model is to some extent data-driven, to preserve spatial accuracy (being a key purpose of the model).

Details on variable selection, datasets used, spatial statistics and methods used, scale, accuracy and error margins are discussed below.

Tree model

Datasets are fed into a tree-based model. A tree-based approach underpins how all indicators and variables are structured within our model. This structure then guides the method and order in which the final index is calculated. Essentially, environmental measures as branches in the model are grouped to represent the four domains, which funnel into the final composite environmental quality measure: the 'Green Poverty Index'. Figure 1 above shows the key themes covered by the model and figure 2 below shows the tree structure of the input data making up comprising the themes.

The tree-based approach is flexible. The structure of the tree can be changed to reflect new data availability, and extra datasets or themes can be added.

Currently each theme or 'branch' on the tree is assumed equally important with no weighting applied. In further iterations of the model the themes and/or datasets within the themes can be weighted if they are considered more impactful.

The units of measure within the model vary; they have been standardised using a z score function before summing together. Whilst values are changed and expressed as standard deviations of the input frequency distributions, the relative spatial variations and patterns are maintained. By standardising the values of cells we equalise the impact of 'branch nodes' so that no single input dataset dominates in a theme.

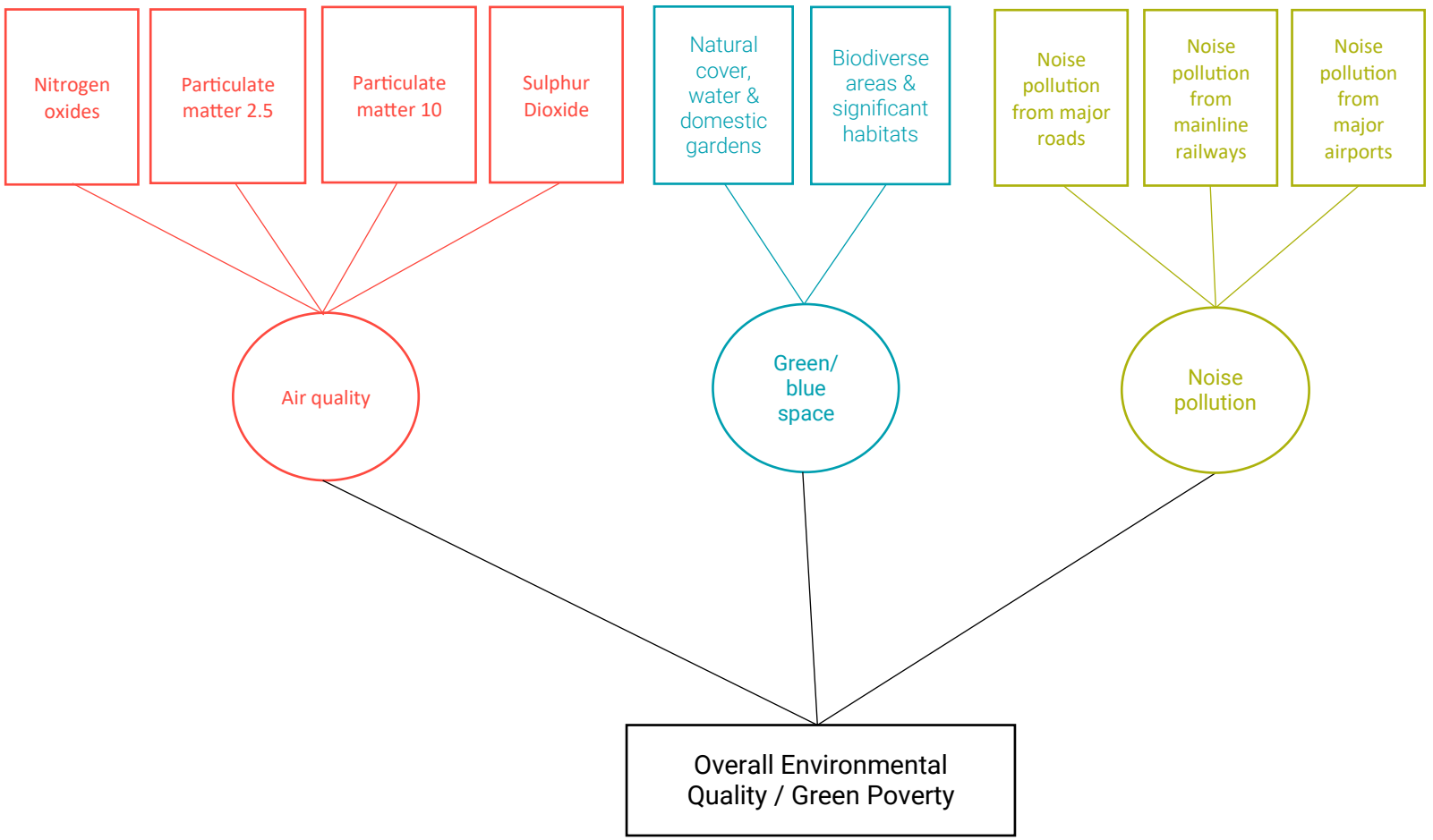
The calculation applied for the standardised z scores is:

$$\frac{(a-b)}{c}$$

a = data point or cell value
b = mean of data points or cell values
c = standard deviation of data points or cell values

The direction of the overall index score is that high values = worse environmental quality / green poverty and low values = better environmental quality / green poverty. Some of the input variable z scores have been reverse-scaled, in order for the direction of each input variable to be logically consistent.

Figure 2: Tree-based model for the Green Poverty Index



Datasets and metrics used

Air quality

- Nitrogen oxides (ug.m-3 NOx as NO2) emissions (estimated) annual mean concentration from all sources, 2020, by 1x1km grid points.
- Particulate matter 2.5 (ug.m-3 gravimetric PM2.5) emissions (estimated) annual mean concentration from all sources, 2020, by 1x1km grid points.
- Particulate matter 10 (ug.m-3 gravimetric PM10) emissions (estimated) annual mean concentration from all sources, 2020, by 1x1km grid points.
- Sulphur dioxide (ug m-3 SO2) emissions (estimated) annual mean concentration from all sources, 2020, by 1x1km grid points.

Data source: UK AIR modelled background pollutant data, 2024

The government National Atmospheric Emission Inventory was considered as it provides modelled emissions data at the small area. However the government 'UK AIR modelled background pollutant data' was used which uses the emissions data with additional modelling to account for weather and atmospheric chemistry, modelling ambient air concentrations which more closely represent the exposure and burden to human health. Z scores were calculated for each of the above measures falling within each grid square. The 4 measures where weighted before being summed together according to their relative impact on populations. The weighted levels were derived from the UK Air Quality Index 2011 calculations[1] ten-point scale of equivalent ug.m-3 mean levels per index score (measuring health effects on the population). This equates to weighting within the air quality theme of nitrogen oxides = 37.7, particulate matter 2.5 = 5.1, particulate matter 10 = 7.1, and sulphur dioxide = 50.

Green/blue space

- Square metres of 'green space', 'water' and 'domestic gardens'
  Data source: Generalised Land Use Database (GLUD) 2010.

- Square metres of 'biodiverse or significant habitat areas'

  Data sources: Special Areas of Conservation (SAC) at Feb 2023, Local Nature Reserves (LNR) at March 2023, National Nature Reserves (NNR) at April 2023, Sites of Special Scientific Interest (SSSI) at April 2023, Priority Habitat Inventory (PHI) at Dec 2022, Wood-Pasture and Parkland BAP Priority Habitat Inventory at Feb 2022, Open mosaic habitats on previously developed land (OMHPDL) at May 2022, Priority River Habitat at Feb 2022, Ramsar sites at Feb 2023, Special Protection Areas (SPA) at Feb 2023, RSPB reserves at Feb 2023, National Forest Inventory at Nov 2022.

[1] Review of the UK Air Quality Index, Committee on the Medical Effects of Air Pollutants, 2011.

The model captures an overview of natural land cover areas by using the Generalised Land Use Database which uses large-scale detailed mapping (OS Mastermap) to aggregate land cover into census 2001 output areas.

The average area (square metres) of 'green and blue' land cover for each output area was multiplied by the actual area of output areas intersecting the 1km grid squares. Output areas are of different absolute sizes hence the average being used, and they do not tessellate into the grid squares. As such, output areas were intersected with the grid squares to achieve an overall estimate, and an averaging error is assumed, to give a best estimate. We may expect the biggest errors using this method to be larger rural output areas, yet these will also contain a majority of non-urban land cover throughout, limiting this error.

The input GLUD dataset uses census boundaries for the full 'extent of the realm', which in some coastal areas cover the sea/tidal water. The categories included for 'water' as defined by OS Mastermap are: inland water, tidal water / inland water, natural environment / inland water; structure / inland water; road or track / inland water; structure / historic interest; inland water. For 'greenspace' this includes the descriptive groups : natural environment / natural general surface / natural road or track / natural rail / natural environment; road or track / natural environment; rail / historic interest.

Z scores are calculated for each grid square based on these final sums.

The model also uses a series of datasets pertaining to special and protected habitats and areas under designation for plants and wildlife as a best estimate proxy dataset for areas of 'biodiversity' or significant habitat (see above for data sources). Some smaller sites will be missed in this approach, for example woodland under 0.5 hectares, but it provides a good general estimate of green space that is of significant ecological value. Places included in more than one dataset are not double-counted: it is common for some places to be covered by more than one designation and therefore included in more than one dataset.

An area calculation in hectares for the extent of land covered by these 'biodiverse' areas was calculated for each grid square and z scores calculated based on these areas.

Both sets of z scores for greenspace/water/domestic gardens plus biodiverse areas, were added together with equal importance (no weighting applied) and a new z score was calculated.

Noise pollution

- Annual average dB over 24 hours from major road traffic routes within areas on all motorways and A class roads with a population of at least 100,000 people (agglomerations) and elsewhere sections of road which have a bi-directional flow of 3 million vehicle passages or more per year (with separate weightings for the evening +5db and night +10db periods), 2017 – capturing places experiencing over 55 dB.

- Annual average dB over 24 hours from all mainline railways within areas with a population of at least 100,000 people (agglomerations) and elsewhere sections of rail route above a flow threshold of 30,000 vehicle passages per year (with separate weightings for the evening and night periods), 2017 – capturing places experiencing over 55 dB.

- Annual average dB over 24 hours of A-weighted continuous sound pressure level which is a notional continuous level from major airports – evening values (1900 – 2300) weighted by the addition of 5 dB (A) and night values (2300 – 0700) weighted by the addition of 10 dB (A) – 2016 - capturing places experiencing over 55 dB.

Data source: Environmental Noise Directive (END) Noise Mapping Agglomerations England Round 3 at 2017, Airport Noise Action Plans (NAPs) with noise modelling at 2016.

For road and rail noise, contours of the dataset are broken down into bands of 55.0-59.9 / 60.0-64.9 / 65.0-69.9 / 70.0-74.9 / >=75.0. For each of these ranges a numeric mid-point has been used as: 57.45 / 62.45 / 67.45 / 72.45 / 77.45 respectively.

For aircraft noise, contours of the dataset are broken down into bands of 55-60 / 60-65 / 65-70 / 70-75 / 75 or more. For each of these ranges a numerical mid-point has been used as: 57.5 / 62.5 / 67.5 / 72.5 / 77.5 respectively.

The area (square metres) of each noise contour falling within each grid square was multiplied by this dB average value, and these values summed for each grid square, before z scores were calculated for each grid square from these final sums.

Spatial data for airport noise has been captured from georeferenced maps at a scale of between 1:50,000 and 1:250,000. A small error margin is assumed in capturing this secondary data.

Some data is missing. Smaller minor roads, and some airports that were not required by government to produce a Noise Action Plan (NAP). Also some smaller roads will have noise hotspots which has not been picked up by the data in this model. In the case of airports, requirements for noise modelling to be produced for a NAP is only in airport areas where there is a resident population in excess of 100,000 persons and a population density equal to or greater than 500 people per km2.[2] Airports such as Exeter, whilst producing significant noise pollution, affect smaller populations out-with of the city and data is missing from the model in these cases. The airports included are: Birmingham, Bournemouth, Bristol, Cambridge, East Midlands, Leeds Bradford, Liverpool, London City, London Gatwick, London Heathrow, London Stansted, Luton, Manchester, Newcastle, Southampton, Southend, and Wycombe Air Park.

Road, rail and airport noise are weighted equally in the model. Noise from industry sources has not been included due to a lack of small-area data, but could be added in at a future date if good data could be sourced.

[2] Department for Environment, Food and Rural Affairs, Guidance for Airport Operators to produce noise action plans under the terms of the Environment Noise (England) Regulations 2006 (as amended), July 2013.

Datasets used

Theme	Measures	Data source	Data owner	Date	Scale
Air quality	Emissions of nitrogen oxides, particulate matter 2.5, particulate matter 10, & sulphur dioxide	National Atmospheric Emissions Inventory	Department for Business, Energy & Industrial Strategy	2022	1 hectare grid square
Green/blue space	Area of 'green space', 'water' and 'domestic gardens'	Generalised Land Use Database	Ministry of Housing, Communities & Local Government	2010	Output Area 2001
	Habitats and species considered the most important at the European level	Special Areas of Conservation (SAC)	Natural England	Feb 2023	1:1,250 scale in urban areas, 1:2,500 in rural areas and 1:10,000 in remote areas
	Statutory designated places with wildlife or geological features that are of special interest locally	Local Nature Reserves (LNR)	Natural England	March 2023	1:1,250 scale in urban areas, 1:2,500 in rural areas and 1:10,000 in remote areas
	Statutory designated places of habitat and geology, including most types of habitats found in England	National Nature Reserves (NNR)	Natural England	April 2023	1:1,250 scale in urban areas, 1:2,500 in rural areas and 1:10,000 in remote areas
	Statutory designated finest sites for wildlife and natural features in England	Sites of Special Scientific Interest (SSSI)	Natural England	April 2023	1:1,250 scale in urban areas, 1:2,500 in rural areas and 1:10,000 in remote areas
	Habitats of principal importance for conservation and biodiversity	Priority Habitat Inventory (PHI)	Natural England	Dec 2022	1:1,250 scale in urban areas, 1:2,500 in rural areas and 1:10,000 in remote areas

	Habitat of principle importance, not included in the PHI dataset	Wood-Pasture and Parkland BAP Priority Habitat Inventory	Natural England	Feb 2022	Various
	Habitat of principle importance, not included in the PHI dataset	Open mosaic habitats on previously developed land (OMHPDL)	Natural England	May 2022	1:1,250 scale in urban areas, 1:2,500 in rural areas and 1:10,000 in remote areas
	Rivers that exhibit a high degree of 'naturalness' which support characteristic and priority species	Priority River Habitat	Natural England	Feb 2022	1:25,000
	Wetlands of International Importance designated via an international UNESCO treaty	Ramsar sites	Natural England	Feb 2023	1:1,250 scale in urban areas, 1:2,500 in rural areas and 1:10,000 in remote areas
	Statutory protected sites for rare and vulnerable birds	Special Protection Areas (SPA)	Natural England	Feb 2023	1:1,250 scale in urban areas, 1:2,500 in rural areas and 1:10,000 in remote areas
	Reserves for important bird habitats	RSPB reserves	Royal Society for the Protection of Birds (RSPB)	Feb 2023	1:10,000
	Forestry and tree cover above 0.5 hectares with a minimum of 20% canopy cover	National Forest Inventory	Forestry Commission	Nov 2022	1:1250, 1:2500 and 1:10000 survey data
Noise pollution	Annual average dB over 24 hours from major road traffic and mainline railways	Strategic noise mapping for England	DEFRA	2017	Modelled on a 10m grid

	Annual average dB over 24 hours from major airports	Airport Noise Action Plans (NAPs) with noise contour maps	Individual airport website reports	2016	Unknown

Interpreting the results

Low scores show better overall environmental quality, higher scores show worse overall environmental quality.

The results are a relative measure of environmental quality or 'green poverty' as compared to other 1km x 1km square locations in England. If the model was calculated on a smaller or larger geographic area, the scores for each grid square would change to reflect the relative score to a set of grid squares within a different study area. The metric used to measure green poverty is therefore not absolute. Each of the 3 themed inputs are weighted equally – if more importance is given to any one of these themes (e.g. air pollution), the scores would also change.

It is a combined measure, so the areas showing as worst may not be suffering from the worst conditions of any one theme (e.g. air pollution or green/blue space), but a location where all types of poor environmental quality was present to some degree.

Spatial models often suffer from the Modifiable Area Unit Problem (MAUP), which is a statistical error introduced when comparing aggregated statistics by different sizes and shapes. We have reduced this error by using relatively small sized grids of equal size. However the spatial error is not non-existent: natural features do not follow grid squares and so to best interpret the result of any one grid square, it is important to look at the patterns and scores in the surrounding area to set this score into a wider context and pattern. Ask the question: is a particularly low or high score the same across the surrounding grid squares or wider area, or limited to that location alone. Examining the results against a map of the input datasets is also useful to make sense of what factors are driving the scores.

For interest, help or use of the data, get in touch.