Supplementary Materials for the manuscript entitled

A Framework for Scalable Ambient Air Pollution Concentration Estimation

S1 Data Details

This section details the preprocessing and transformations to create a consistent dataset for training the data-driven supervised machine learning model.

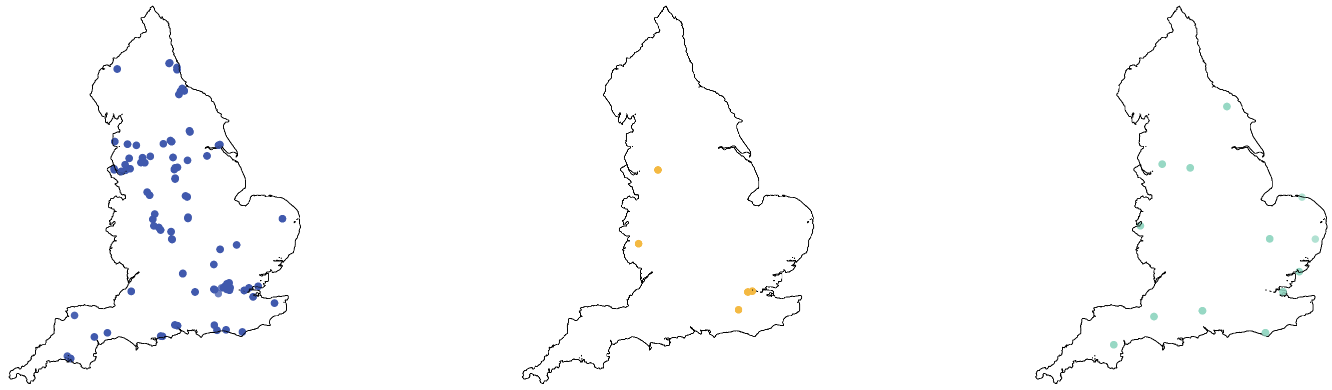
S1.1 Common Data Format

As the model framework is Eulerian [1], the first decision was the grid framework, taking into context the modifiable areal unit problem (MAUP) [2] for the grids in which the aggregation for predictors would be taken. The decision was made to use the same framework as existing air pollution concentration datasets, particularly the UK Modelled Background Annual dataset [3].

Supplementary Figure S1: **1km^2 land grids For England.** The land grids covering the area of the England land mass provide the common framework for aggregating the datasets and providing estimates of ambient air pollution, a total of 355,827 point locations at the centroid of each grid for which measurements are sampled.

S1.2 Air Pollution Concentrations

Detailed are the supporting analysis and figures for the air pollution concentration data. Included are the spatial distribution of the AURN network monitoring stations for each top-level environmental classification in Figure S2, kernel density estimates for each air pollutant concentration dataset Figure S3, and the abstracted distance of each AURN monitoring station from its real location in the model framework Table S3.



(a) 91 Unique Urban Station Locations

(b) 5 Unique Suburban Station Locations

(c) 13 Unique Rural Station Locations

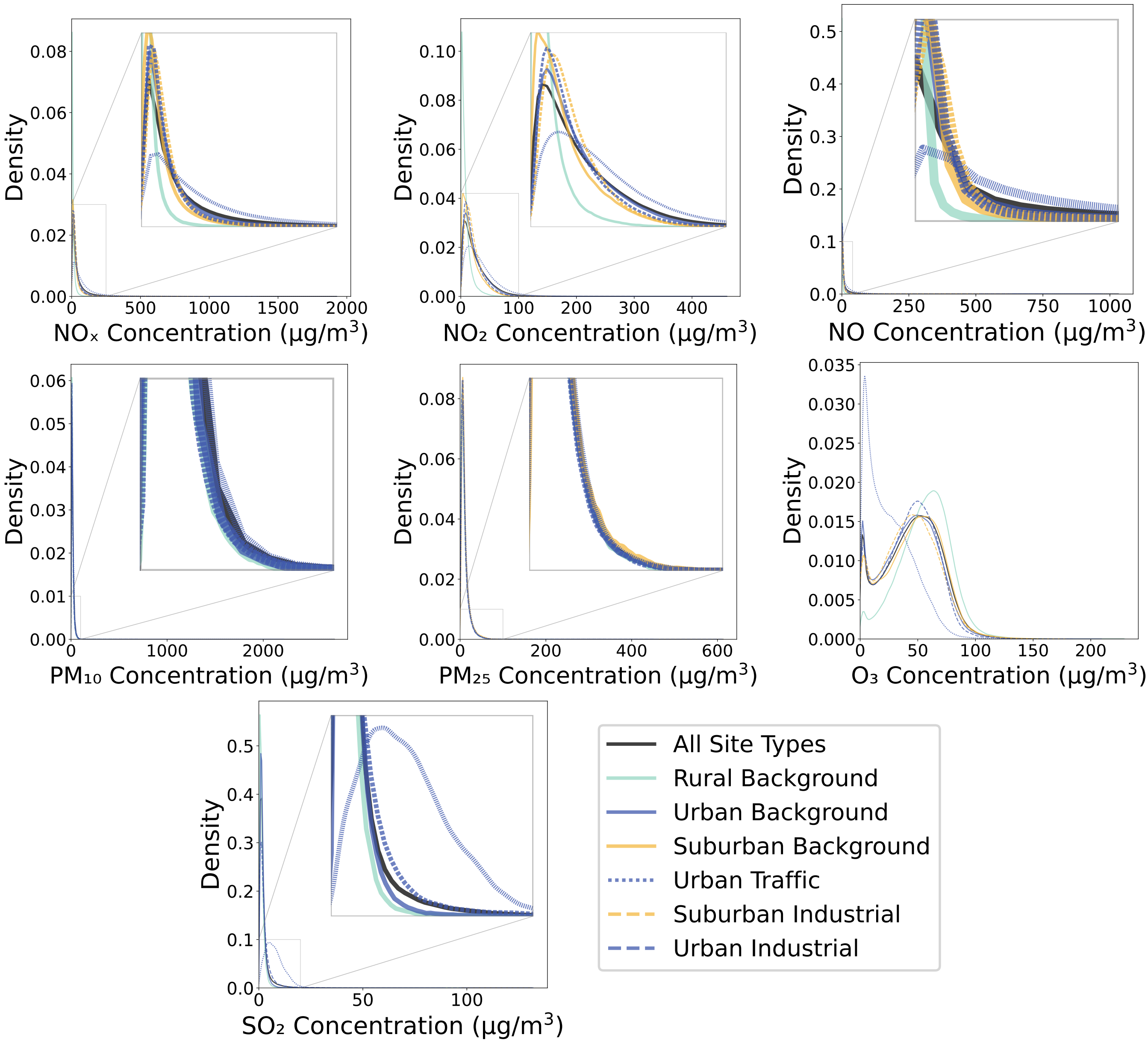
Supplementary Figure S2: **Spatial distribution and classification of monitoring stations within England.** The Automatic Urban and Rural Network (AURN) stations are divided into 3 classes: urban, suburban, and rural. The urban stations are then further divided into background, traffic and industrial, suburban into background and industrial and rural background. Note the inequality of station numbers; most of the stations are in urban settings.

AURN Site Name	Peak Value Timestamp	Peak Value (μg/m^3)	Peak Day Average (μg/m^3)	Peak Year Average (μg/m^3)
London Marylebone Road	10/06/2016 15:00:00	321.9	139.7	89.1
Sandy Roadside	01/03/2018 16:00:00	336.7	43.2	26.6
Luton A505 Roadside	19/01/2016 08:00:00	362.7	176.9	49.8
Camden Kerbside	15/02/2016 18:00:00	385.3	124.4	65.9
Manchester Piccadilly	13/09/2014 15:00:00	456.8	69.9	40.5

Supplementary Table S1: **AURN NO_2 monitoring station peak values between 2014-2018 with associated year daily mean peak and annual mean; for the year in which the peak measurement occurs.** These five monitoring stations show how there is not a simple relationship between the peak value, the peak daily average and the overall peak year average. London Marylebone Road station never has a peak as intense as Manchester Piccadilly but does experience consistently higher pollution across the year and similarly has a higher peak daily mean. However, Sandy Roadside has a higher overall peak than London Marylebone Road but a considerably lower peak day and year mean. The five stations have highlighted how multiple averages and a finer temporal scale are needed to uncover the intricacies of air pollution experienced at a single location.

Pollutant Name	Number Of Values	Number of Negative Values	Percentage of Negative Values (%)
NO_x	3855871	235	0.01
NO_2	3855760	2697	0.07
NO	3859339	2078	0.05
O_3	2198186	659	0.03
PM_{10}	1646086	3743	0.23
PM_{25}	2004487	23,515	1.17
SO_2	683276	1016	0.15

Supplementary Table S2: **AURN negative data point summary.** Negative data points within the AURN air pollution concentrations were removed from the dataset as the only form of preprocessing performed on the dataset. Negative concentrations can't exist, and their presence in the dataset indicates a fault with the instruments at the monitoring station.



Supplementary Figure S3: **Kernel Density Estimation (KDE) for air pollution measurements.** The KDE plots for each air pollutant show the distribution of concentration values. For PM_{10} and $PM_{2.5}$, the distribution of values appears to remain constant across all sub classifications of the monitoring station, whereas NO_x, NO_2, NO and SO_2 appear to have different concentrations at different environmental locations. All air pollutants other than O_3 have a right skew, indicating a tendency for no air pollution to be the norm in the atmosphere, with pollution emitted and then subsequently dispersed, reducing the concentration measured.

AURN Station Name	Station Environment Type	Station Latitude	Station Longitude	Grid Centroid Latitude	Grid Centroid Longitude	Station Distance From Grid Centroid (m)
Plymouth Tavistock Road	Urban Traffic	50.4	-4.1	50.4	-4.1	8.7
Aston Hill	Rural Background	52.5	-3.0	52.5	-3.0	18.2
Salford Eccles	Urban Background	53.5	-2.3	53.5	-2.3	27.4
Eastbourne	Urban Background	50.8	0.3	50.8	0.3	31.7
Leamington Spa Rugby Road	Urban Traffic	52.3	-1.5	52.3	-1.5	32.4
Birkenhead Borough Road	Urban Traffic	53.4	-3.0	53.4	-3.0	35.0
Blackpool Marton	Urban Background	53.8	-3.0	53.8	-3.0	35.1
Wharleycroft	Rural Background	54.6	-2.5	54.6	-2.5	38.1
London Brent	Urban Background	51.6	-0.3	51.6	-0.3	39.6
London Bridge Place	Urban Background	51.5	-0.1	51.5	-0.1	46.7
Liverpool Queen's Drive Roadside	Urban Traffic	53.4	-3.0	53.4	-3.0	47.1
Lullington Heath	Rural Background	50.8	0.2	50.8	0.2	54.1
Great Dun Fell	Rural Background	54.7	-2.5	54.7	-2.5	59.1
Birmingham A4540 Roadside	Urban Traffic	52.5	-1.9	52.5	-1.9	59.2
Scunthorpe Town	Urban Industrial	53.6	-0.6	53.6	-0.6	63.4
Rotherham Centre	Urban Background	53.4	-1.4	53.4	-1.4	65.8
London Westminster	Urban Background	51.5	-0.1	51.5	-0.1	67.1
Wigan Centre	Urban Background	53.5	-2.6	53.5	-2.6	68.3
Tower Hamlets Roadside	Urban Traffic	51.5	-0.0	51.5	-0.0	70.3
Middlesbrough	Urban Industrial	54.6	-1.2	54.6	-1.2	72.0
Immingham Woodlands Avenue	Urban Background	53.6	-0.2	53.6	-0.2	73.1
Billingham	Urban Industrial	54.6	-1.3	54.6	-1.3	87.2
London Bromley	Urban Traffic	51.4	0.0	51.4	0.0	90.8
Bury Roadside	Urban Traffic	53.5	-2.3	53.5	-2.3	91.1
Cambridge	Urban Traffic	52.0	0.0	52.0	0.0	96.4
London Teddington	Urban Background	51.4	-0.3	51.4	-0.3	96.8
London Cromwell Road 2	Urban Traffic	51.5	-0.2	51.5	-0.2	98.3
Reading London Road	Urban Traffic	51.5	-0.9	51.5	-0.9	98.5
Leeds Potternewton	Urban Background	53.8	-1.5	53.8	-1.5	98.8
London Haringey Priory Park South	Urban Background	51.6	-0.1	51.6	-0.1	100.3
Birmingham East	Urban Background	52.5	-1.8	52.5	-1.8	103.7
Birmingham Acocks Green	Urban Background	52.4	-1.8	52.4	-1.8	111.4
Charlton Mackrell	Rural Background	51.1	-2.7	51.1	-2.7	112.6
Northampton	Urban Background	52.3	-0.9	52.3	-0.9	114.1
Saltash Callington Road	Urban Traffic	50.4	-4.2	50.4	-4.2	114.1
Bottesford	Rural Background	52.9	-0.8	52.9	-0.8	114.4
Leamington Spa	Urban Background	52.3	-1.5	52.3	-1.5	116.0
Hove Roadside	Urban Traffic	50.8	-0.2	50.8	-0.2	119.5
Oxford Centre Roadside	Urban Traffic	51.8	-1.3	51.8	-1.3	120.5
Walsall Alumwell	Urban Background	52.6	-2.0	52.6	-2.0	121.3
Sandwell Oldbury	Urban Background	52.5	-2.0	52.5	-2.0	124.0
Coventry Allesley	Urban Background	52.4	-1.6	52.4	-1.6	124.1
Sunderland Wessington Way	Urban Traffic	54.9	-1.4	54.9	-1.4	125.6
Manchester Town Hall	Urban Background	53.5	-2.2	53.5	-2.2	128.6
Sunderland	Urban Background	54.9	-1.4	54.9	-1.4	128.9
Sheffield Centre	Urban Background	53.4	-1.5	53.4	-1.5	129.6
Bristol Old Market	Urban Traffic	51.5	-2.6	51.5	-2.6	133.3
Hartlepool St Abbs Walk	Urban Background	54.7	-1.2	54.7	-1.2	137.5
Southampton A33	Urban Traffic	50.9	-1.5	50.9	-1.5	137.5
Liverpool Speke	Urban Industrial	53.3	-2.8	53.3	-2.8	145.6
Bradford Centre	Urban Background	53.8	-1.7	53.8	-1.7	145.8
Doncaster A630 Cleveland Street	Urban Traffic	53.5	-1.1	53.5	-1.1	148.5
London Bloomsbury	Urban Background	51.5	-0.1	51.5	-0.1	151.1
London Southwark	Urban Background	51.5	-0.1	51.5	-0.1	151.2
Liverpool Centre	Urban Background	53.4	-3.0	53.4	-3.0	152.3
Thurrock	Urban Background	51.5	0.3	51.5	0.3	154.7
Chesterfield Loundsley Green	Urban Background	53.2	-1.5	53.2	-1.5	155.2
Northampton Kingsthorpe	Urban Background	52.3	-0.9	52.3	-0.9	162.9
Derby St Alkmund's Way	Urban Traffic	52.9	-1.5	52.9	-1.5	165.7
Brentford Roadside	Urban Traffic	51.5	-0.3	51.5	-0.3	169.5
Birmingham Ladywood	Urban Background	52.5	-1.9	52.5	-1.9	169.8
Sheffield Barnsley Road	Urban Traffic	53.4	-1.5	53.4	-1.5	170.0
Market Harborough	Rural Background	52.6	-0.8	52.6	-0.8	172.2
Northampton Spring Park	Urban Background	52.3	-0.9	52.3	-0.9	174.4
Canterbury	Urban Background	51.3	1.1	51.3	1.1	179.1
Redcar	Suburban Background	54.6	-1.1	54.6	-1.1	180.7
Burton-on-Trent Horninglow	Urban Background	52.8	-1.6	52.8	-1.6	182.9
Bristol Centre	Urban Background	51.5	-2.6	51.5	-2.6	183.9
Ladybower	Rural Background	53.4	-1.8	53.4	-1.7	186.5
Norwich Lakenfields	Urban Background	52.6	1.3	52.6	1.3	187.6
Sheffield Devonshire Green	Urban Background	53.4	-1.5	53.4	-1.5	188.2
Blackburn Accrington Road	Urban Traffic	53.7	-2.5	53.7	-2.5	191.7
London Cromwell Road	Urban Traffic	51.5	-0.2	51.5	-0.2	192.5
Wirral Tranmere	Urban Background	53.4	-3.0	53.4	-3.0	194.8
Sibton	Rural Background	52.3	1.5	52.3	1.5	195.1
Bristol St Paul's	Urban Background	51.5	-2.6	51.5	-2.6	195.4
Lincoln Canwick Road	Urban Traffic	53.2	-0.5	53.2	-0.5	195.6
Weybourne	Rural Background	53.0	1.1	53.0	1.1	196.5
Carlisle Roadside	Urban Traffic	54.9	-2.9	54.9	-2.9	196.9
Blackburn Darwen Roadside	Urban Traffic	53.7	-2.5	53.7	-2.5	197.9
London N. Kensington	Urban Background	51.5	-0.2	51.5	-0.2	199.4
Southwark Roadside	Urban Traffic	51.5	-0.1	51.5	-0.1	199.4
Stockport Shaw Heath	Urban Background	53.4	-2.2	53.4	-2.2	201.0
Stanford-le-Hope Roadside	Urban Traffic	51.5	0.4	51.5	0.4	201.8
High Muffles	Rural Background	54.3	-0.8	54.3	-0.8	204.0
Camden Kerbside	Urban Traffic	51.5	-0.2	51.5	-0.2	209.0
Leominster	Suburban Background	52.2	-2.7	52.2	-2.7	209.5
York Fishergate	Urban Traffic	54.0	-1.1	54.0	-1.1	209.5
Leicester University	Urban Background	52.6	-1.1	52.6	-1.1	210.4
Honiton	Urban Background	50.8	-3.2	50.8	-3.2	211.2
West Bromwich Kenrick Park	Urban Background	52.5	-2.0	52.5	-2.0	211.4
Sandwell West Bromwich	Urban Background	52.5	-2.0	52.5	-2.0	216.6
Hull Centre	Urban Background	53.7	-0.3	53.7	-0.3	217.1
Oldbury Birmingham Road	Urban Traffic	52.5	-2.0	52.5	-2.0	217.7
Leeds Headingley Kerbside	Urban Traffic	53.8	-1.6	53.8	-1.6	222.4
Nottingham Centre	Urban Background	53.0	-1.1	53.0	-1.1	224.5
Wigan Leigh	Urban Background	53.5	-2.5	53.5	-2.5	224.5
Lincoln Roadside	Urban Traffic	53.2	-0.5	53.2	-0.5	225.3
Southend-on-Sea	Urban Background	51.5	0.7	51.5	0.7	225.5
Chesterfield	Urban Background	53.2	-1.4	53.2	-1.4	226.9
Coventry Memorial Park	Urban Background	52.4	-1.5	52.4	-1.5	228.1
Plymouth Centre	Urban Background	50.4	-4.1	50.4	-4.1	228.9
Bradford Mayo Avenue	Urban Traffic	53.8	-1.8	53.8	-1.8	229.6
Swindon Walcot	Urban Background	51.6	-1.8	51.6	-1.8	232.0
Rugeley	Urban Background	52.8	-1.9	52.8	-1.9	232.8
Warrington	Urban Industrial	53.4	-2.6	53.4	-2.6	233.0
London Harlington	Urban Industrial	51.5	-0.4	51.5	-0.4	233.9
Bromley Roadside	Urban Traffic	51.4	0.0	51.4	0.0	234.1

(a) **AURN station real distance from the centroid distance.**

AURN Station Name	Station Environment Type	Station Latitude	Station Longitude	Grid Centroid Latitude	Grid Centroid Longitude	Station Distance From Grid Centroid (m)
London Honor Oak Park	Urban Background	51.4	-0.0	51.4	-0.0	234.5
Saltash Roadside	Urban Traffic	50.4	-4.2	50.4	-4.2	234.5
Ealing Horn Lane	Urban Traffic	51.5	-0.3	51.5	-0.3	236.0
Bolton	Urban Background	53.6	-2.4	53.6	-2.4	237.4
Hull Holderness Road	Urban Traffic	53.8	-0.3	53.8	-0.3	238.3
Sheffield Tinsley	Urban Background	53.4	-1.4	53.4	-1.4	238.6
Widnes Milton Road	Urban Traffic	53.4	-2.7	53.4	-2.7	240.2
London Bexley	Suburban Background	51.5	0.2	51.5	0.2	240.7
Newcastle Cradlewell Roadside	Urban Traffic	55.0	-1.6	55.0	-1.6	242.3
Cambridge Roadside	Urban Traffic	52.2	0.1	52.2	0.1	243.8
Central London	Urban Background	51.5	-0.1	51.5	-0.1	244.1
Chesterfield Roadside	Urban Traffic	53.2	-1.5	53.2	-1.5	244.9
Manchester South	Suburban Industrial	53.4	-2.2	53.4	-2.2	245.9
Haringey Roadside	Urban Traffic	51.6	-0.1	51.6	-0.1	246.6
Stoke-on-Trent Centre	Urban Background	53.0	-2.2	53.0	-2.2	246.6
Luton A505 Roadside	Urban Traffic	51.9	-0.5	51.9	-0.5	248.1
Chilworth	Suburban Background	51.0	-1.4	51.0	-1.4	248.9
Norwich Centre	Urban Background	52.6	1.3	52.6	1.3	249.2
Hounslow Roadside	Urban Traffic	51.5	-0.3	51.5	-0.3	251.6
London Haringey	Urban Background	51.6	-0.1	51.6	-0.1	251.9
Featherstone	Urban Background	53.7	-1.4	53.7	-1.4	255.0
York Bootham	Urban Background	54.0	-1.1	54.0	-1.1	256.3
Stockton-on-Tees Eaglescliffe	Urban Traffic	54.5	-1.4	54.5	-1.4	256.5
London Harrow	Suburban Background	51.6	-0.4	51.6	-0.3	257.7
Wicken Fen	Rural Background	52.3	0.3	52.3	0.3	258.0
Birmingham Tyburn	Urban Background	52.5	-1.8	52.5	-1.8	259.8
Bristol East	Urban Background	51.5	-2.6	51.5	-2.6	259.8
Scunthorpe	Urban Industrial	53.6	-0.6	53.6	-0.6	260.5
Leeds Centre	Urban Background	53.8	-1.5	53.8	-1.5	261.7
Southampton Centre	Urban Background	50.9	-1.4	50.9	-1.4	262.4
Stewartby	Urban Industrial	52.1	-0.5	52.1	-0.5	262.7
London Marylebone Road	Urban Traffic	51.5	-0.2	51.5	-0.2	263.1
Brighton Preston Park	Urban Background	50.8	-0.1	50.8	-0.2	263.1
Leicester Centre	Urban Background	52.6	-1.1	52.6	-1.1	265.6
Hull Freetown	Urban Background	53.7	-0.3	53.7	-0.3	266.2
Borehamwood Meadow Park	Urban Background	51.7	-0.3	51.7	-0.3	267.8
Dewsbury Ashworth Grove	Urban Background	53.7	-1.6	53.7	-1.6	267.9
Wolverhampton Centre	Urban Background	52.6	-2.1	52.6	-2.1	268.2
Stockton-on-Tees Yarm	Urban Traffic	54.5	-1.4	54.5	-1.4	270.0
Stockport	Urban Background	53.4	-2.2	53.4	-2.2	273.3
Birmingham Kerbside	Urban Traffic	52.3	-1.9	52.3	-1.9	273.7
Shaw Crompton Way	Urban Traffic	53.6	-2.1	53.6	-2.1	275.1
Telford Hollinswood	Urban Background	52.7	-2.4	52.7	-2.4	278.6
Bury Whitefield Roadside	Urban Traffic	53.6	-2.3	53.6	-2.3	281.6
Walsall Willenhall	Urban Background	52.6	-2.0	52.6	-2.0	282.4
Stockton-on-Tees A1305 Roadside	Urban Traffic	54.6	-1.3	54.6	-1.3	282.7
Chilbolton Observatory	Rural Background	51.1	-1.4	51.2	-1.4	283.5
Wray	Rural Background	54.1	-2.6	54.1	-2.6	283.8
Bristol Temple Way	Urban Traffic	51.5	-2.6	51.5	-2.6	284.6
Sutton Roadside	Urban Traffic	51.4	-0.2	51.4	-0.2	287.4
Blackpool	Urban Background	53.8	-3.0	53.8	-3.0	288.6
Yarner Wood	Rural Background	50.6	-3.7	50.6	-3.7	288.8
Exeter Roadside	Urban Traffic	50.7	-3.5	50.7	-3.5	289.2
London UCL	Urban Background	51.5	-0.1	51.5	-0.1	290.5
Harwell	Rural Background	51.6	-1.3	51.6	-1.3	290.6
Coventry Binley Road	Urban Traffic	52.4	-1.5	52.4	-1.5	292.1
St Helens Linkway	Urban Traffic	53.5	-2.7	53.5	-2.7	292.3
Horley	Suburban Industrial	51.2	-0.2	51.2	-0.2	293.0
Barnstaple A39	Urban Traffic	51.1	-4.0	51.1	-4.0	293.2
Manchester Piccadilly	Urban Background	53.5	-2.2	53.5	-2.2	295.1
Walsall Woodlands	Urban Background	52.6	-2.0	52.6	-2.0	297.2
Southwark A2 Old Kent Road	Urban Traffic	51.5	-0.1	51.5	-0.1	297.4
Birmingham Tyburn Roadside	Urban Traffic	52.5	-1.8	52.5	-1.8	297.4
London Islington	Urban Background	51.5	-0.1	51.5	-0.1	298.2
Rochester Stoke	Rural Background	51.5	0.6	51.5	0.6	299.0
London Harrow Stanmore	Urban Background	51.6	-0.3	51.6	-0.3	299.9
Manchester Sharston	Suburban Industrial	53.4	-2.2	53.4	-2.2	302.0
Storrington Roadside	Urban Traffic	50.9	-0.4	50.9	-0.4	302.0
Stevenage	Suburban Background	51.9	-0.2	51.9	-0.2	305.3
Sunderland Silksworth	Urban Background	54.9	-1.4	54.9	-1.4	305.4
Newcastle Centre	Urban Background	55.0	-1.6	55.0	-1.6	305.5
London Lewisham	Urban Background	51.4	-0.0	51.4	-0.0	305.7
Stoke-on-Trent A50 Roadside	Urban Traffic	53.0	-2.1	53.0	-2.1	306.6
London A3 Roadside	Urban Traffic	51.4	-0.3	51.4	-0.3	309.0
Norwich Forum Roadside	Urban Traffic	52.6	1.3	52.6	1.3	310.6
London Teddington Bushy Park	Urban Background	51.4	-0.3	51.4	-0.3	313.7
Leicester A594 Roadside	Urban Traffic	52.6	-1.1	52.6	-1.1	314.9
Barnsley 12	Urban Background	53.6	-1.5	53.6	-1.5	319.3
Chatham Roadside	Urban Traffic	51.4	0.5	51.4	0.5	320.8
West London	Urban Background	51.5	-0.2	51.5	-0.2	323.2
Glazebury	Rural Background	53.5	-2.5	53.5	-2.5	323.4
Barnsley	Urban Background	53.6	-1.5	53.6	-1.5	323.6
Barnsley Gawber	Urban Background	53.6	-1.5	53.6	-1.5	326.0
Brighton Roadside	Urban Traffic	50.8	-0.1	50.8	-0.1	326.4
Crewe Coppenhall	Urban Background	53.1	-2.5	53.1	-2.5	330.3
Reading New Town	Urban Background	51.5	-0.9	51.5	-0.9	330.9
Bircotes	Urban Background	53.4	-1.1	53.4	-1.1	332.9
Worthing A27 Roadside	Urban Traffic	50.8	-0.4	50.8	-0.4	335.7
Cannock A5190 Roadside	Urban Traffic	52.7	-2.0	52.7	-2.0	336.5
Bath Roadside	Urban Traffic	51.4	-2.4	51.4	-2.4	338.6
London Sutton	Suburban Background	51.4	-0.2	51.4	-0.2	338.8
St Osyth	Rural Background	51.8	1.0	51.8	1.1	340.3
Birmingham Centre	Urban Background	52.5	-1.9	52.5	-1.9	342.6
London Wandsworth	Urban Background	51.5	-0.2	51.5	-0.2	342.8
Christchurch Barrack Road	Urban Traffic	50.7	-1.8	50.7	-1.8	346.4
London Hillingdon	Urban Background	51.5	-0.5	51.5	-0.5	347.8
Somerton	Rural Background	51.0	-2.7	51.0	-2.7	347.8
Nottingham Western Boulevard	Urban Traffic	53.0	-1.2	53.0	-1.2	350.0
Preston	Urban Background	53.8	-2.7	53.8	-2.7	356.2
Norwich Roadside	Urban Traffic	52.6	1.3	52.6	1.3	356.6
Oxford St Ebbes	Urban Background	51.7	-1.3	51.7	-1.3	357.6
London Eltham	Suburban Background	51.5	0.1	51.4	0.1	359.7
Coventry Centre	Urban Background	52.4	-1.5	52.4	-1.5	363.9
Reading	Urban Background	51.5	-1.0	51.5	-1.0	364.6
Sandy Roadside	Urban Traffic	52.1	-0.3	52.1	-0.3	370.1
Bournemouth	Urban Background	50.7	-1.8	50.7	-1.8	384.3
Bath A4 Roadside	Urban Traffic	51.4	-2.4	51.4	-2.4	385.3
London Hackney	Urban Background	51.6	-0.1	51.6	-0.1	399.8

(b) **AURN station real distance from the centroid distance (cont.).** For each AURN monitoring station used within the study, the station's latitude and longitude are given, alongside the abstracted location of the station within the study, denoted by the grid centroids latitude and longitude. The station distance then gives the difference between the stations true location and the location used within the study.

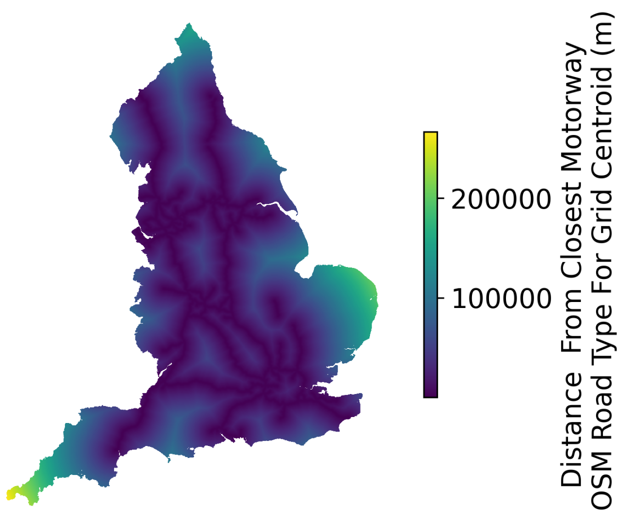
Supplementary Table S3

S1.3 Transport Infrastructure Structural Properties

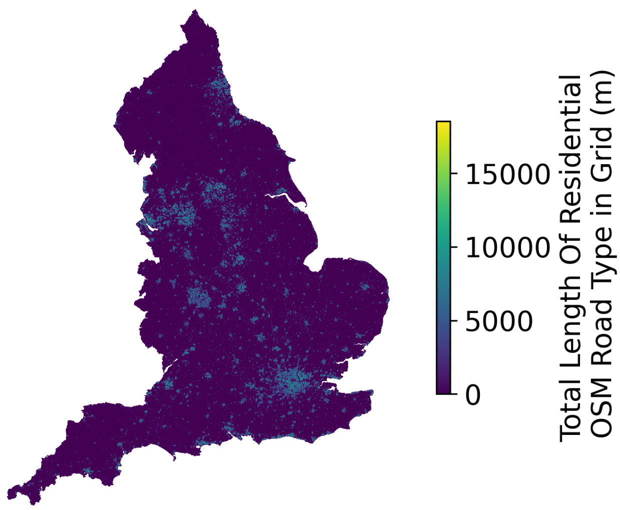
Open Street Maps was used as the data set to build the transport infrastructure feature vector. Open Street Maps provides a high level of detail on the road location and the type of road, alongside providing a historical dataset that allows for historical roads to be acquired across years. Due to the computational cost of retrieving the feature vector for the transport infrastructure in a grided format, and the minimal change to the road infrastructure itself on a fine temporal level, especially hourly, we decided to take yearly snapshots of the road infrastructure. A possible improvement to the method would be to take more frequency snapshots of the road network at the expense of additional computation if desired. The snapshot of the road network used was the road network structural on the first day of the year. We then used this snapshot of the road network to create a feature vector for the following year of timestamps within the feature vector.

The first set of feature vectors concerning transport infrastructure structural properties detailed each grid's distance to the closest road type within the study in meters. Figure S4a shows the feature vector for the distance to the closet motorway in 2018. The second set of feature vectors created concerning transport infrastructure structural properties details the total length in meters of each analysed road type for every grid within the study. Figure S4b shows the total residential road length in meters in all grids for 2018.

The highway types analysed for creating the feature vector for the transport infrastructure structural properties dataset family included Residential, Footway, Service, Primary, Path, Cycleway, Tertiary, Secondary, Unclassified, Trunk, Track, Motorway, Pedestrian and Living Street. Figure S4 provide an example of the full transport infrastructural properties dataset for the distance to the closet motorway and the total length of residential road for each grid.



(a) Grid Centroid Distance To Closest Motorway



(b) Total Length of Residential Road Within Grid

Supplementary Figure S4: **Example complete England transport infrastructure structural properties datasets.** A feature vector element is created for the distance to the closest and the total length of the specific road type. 14 different road types are analysed, resulting in 28 feature vector elements contributed by the transport infrastructure structural properties dataset family.

S1.4 Transport Infrastructure Use

Traffic counts from point locations across England were used to estimate the daily traffic flow across given types of roads across different regions within England. The traffic counts used were part of the Department of Transports (DfT) Road usage data included in the annual average daily flow (AADF), major and minor roads dataset [4]. The AADF dataset estimates a range of different transport methods, with the following aggregate transport types being used in the study:

1. Bicycle Count
2. Car and Taxi Count
3. Bus and Coach Score
4. Light Goods Van (LGV)
5. Heavy Goods Vechile (HGV)

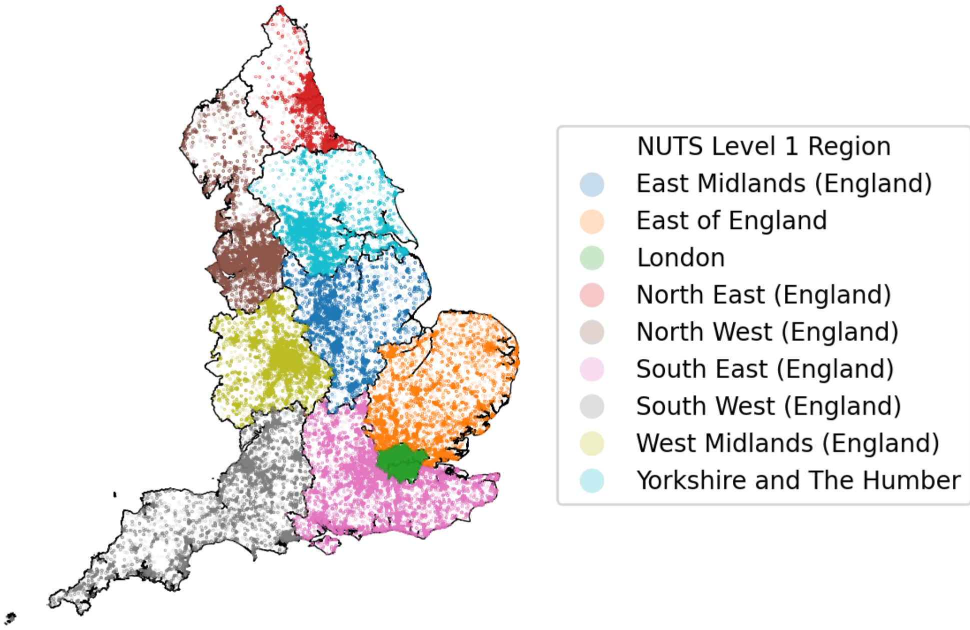
The first step was to create a mean traffic flow per road type. The DfT AADF dataset gives traffic flow estimates on major and minor roads, from motorways to rural areas, including single-lane roads with passing bays [5]. In the transport infrastructural properties dataset family, we included 14 road types, with road types such as cycleways. As the AADF dataset only includes road types suitable for motor vehicles, there was a need to reduce the road types to only those related to the major and minor roads defined by DfT. Therefore, from the OpenStreetMap dataset, we included only the ten road types: motorway, trunk, primary, secondary, tertiary, unclassified, residential, living_street, service, and track. We then matched the sample location from the AADF to the closest OpenStreetMaps road type to calculate a mean for the daily traffic flow for that road.

As road type usage can be substantially different across England, such as a residential road in central London and a small town in the Midlands having widely different traffic flows, we created means for each road type within a set of defined geographic regions. The geographic boundary we chose for this aggregation of sample locations was the NUTS Level 1 Regions, seen in Figure S5. Smaller region sizes of Local Authority Districts, with 371 geographic regions, were also trialled; however, some of the 10 OpenStreetMaps road types had no estimates for the mean traffic flow. Therefore, we used the coarser but more comprehensive aggregation of the NUTS boundary.

The next step was calculating the road network within each grid used within the study. Figure S6 shows the road infrastructure within a single 1km^2 grid in South Cambridgeshire at location Latitude 52.218, Longitude -0.07. We then calculated the total road length for each road type for each grid. Table S4 shows the total road length for each road type for the grid shown in Figure S6 alongside the number of traffic counts within the DfT dataset for that road type within that NUTS 1 region, in this case, the East of England region. Each road type's mean traffic flow per transport method was multiplied by the overall length of that road type to estimate the traffic flow for that transport method across that road type within the 1km^2 grid. Each of the road types multiplication was then summed to provide an overall estimate for the traffic of each transport method across the whole road network within that grid, with Table S5 showing the overall traffic score for each transport method for the grid shown in Figure S6.

The traffic flow score in Table S5 gives an estimate of traffic flow at the daily level. However, an estimate of traffic flow based on an hourly level was desired for the analysis of rush hour traffic. To achieve this, we temporally distributed the daily traffic flow based on a spatial microsimulation of the UK Time Use Survey [6]. The UK Time Use Survey provides data on how 11421 individuals spent their time across the UK during weekdays, Saturdays and Sundays. One of the options for how they could specify how they were spending their time was for travelling, which included details of the transport method they were using. Profiling of travel habits was made possible as each participant has associated socio-demographic data. Using the UK census [7], we used a spatial microsimulation [8] to create a synthetic population of England. The input UK census data was the 7201 Middle Layer Super Output Area (MSOA) aggregate socio-demographic statistics. The spatial microsimulation provided a synthetic population for each MSOA that included data on when they would travel based on the UK Time Use Survey. We then created aggregate travel times for each MSOA region. Figure S7 shows the travel profile for MSOA South Cambridgeshire 020, the MSOA that allowed an understanding of what times of day for a weekday, Saturday and Sunday individuals within a given MSOA travel by transport method. The travel profile seen in Figure S7 was then used to temporally distribute the daily traffic flow score for each grid within that MSOA. The grid shown in Figure S6 is within the NUTS region East of England and MSOA South Cambridgeshire 020. As such, the travel profiles in Figure S7 were used to distribute the daily traffic flows in Table S5 temporally to produce the hourly traffic estimates.

Figure S8 shows the difference in traffic score for grids within a London subset. As seen in the figure, each day exhibits a unique signal, with the weekday seeing the highest travel, followed by Saturday and finally Sunday. Figure S9 shows how the traffic flow within a grid for each transport method differs depending on the road types present within the grid. The figure depicts the most significant traffic flow for HGVs on the arterial lines into London and the M25 ring road around central London. In contrast, bicycle use is most prominent in central London. Figure S10 shows the full transport use dataset across all of the land grids, with the white land grids representing grids within the study area with no roads present. Hence, no traffic flow estimate has been made.



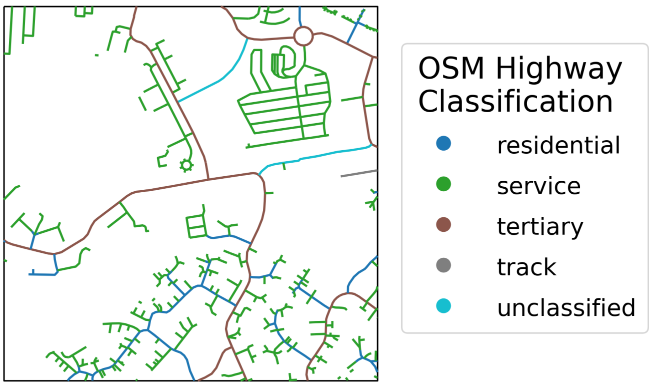
Supplementary Figure S5: **Department for Transport point locations for Annual Average Daily Flow (AADF) of traffic.** Shown are the point sample locations for traffic counts along roads within England that the Department of Transport conducted, giving an annual average daily flow for various transport types. The points are slightly transparent to allow insight into where multiple points overlap, alongside being coloured by corresponding NUTS 1 region, indicating groups of point samples used to calculate an average flow for that region.

Total Road Length (m)	OSM Highway Classification	Traffic Count Counts	Mean Pedal Cycles Traffic	Mean Cars and Taxis Traffic	Mean Bus and Coaches Traffic	Mean LGV Traffic	Mean HGV Traffic
2238.1	residential	448.0	73.2	3954.9	38.9	659.8	129.3
10,138.2	service	180.0	127.2	13,983.9	102.5	2701.8	1033.9
3220.2	tertiary	366.0	60.5	4356.0	33.1	769.7	224.6
99.5	track	58.0	15.8	14,115.8	64.2	3318.3	2036.2
587.6	unclassified	305.0	25.6	3604.1	18.9	742.6	342.6

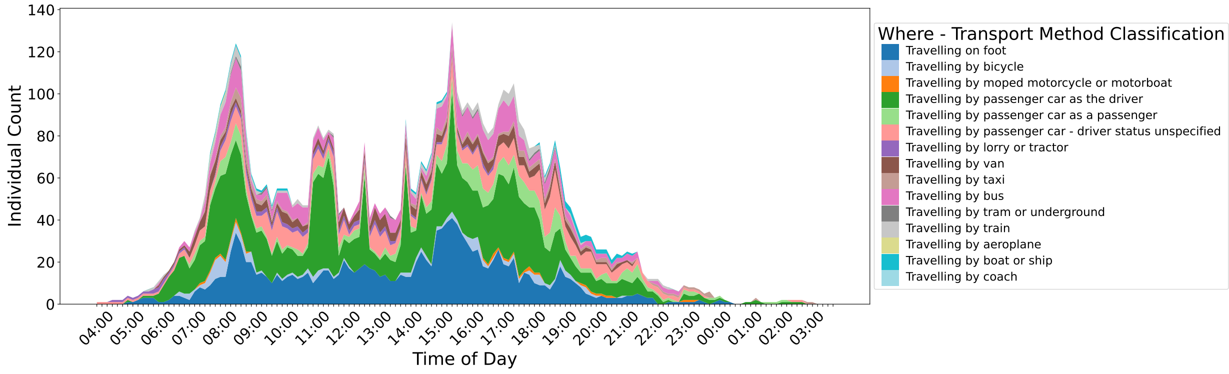
Supplementary Table S4: **Mean traffic count per road for East of England, with summary data road network length for Figure S6** The summary data for the road network within a single grid within the East of England is shown, corresponding to the visualisation in Figure S6. The table provides the data used to create the overall grid score for the grid shown in Table S5. To create the grid score, the sum of the multiplication of the length of the road type and the average traffic flow on that road type gives a single number indicating the traffic for that transport type across the road network within the grid.

OSM Highway Classification	Grid Score
Pedal Cycles	1,664,894
Cars and Taxis	168,171,555
Bus and Coaches	1,250,190
LGV	32,113,165
HGV	11,897,900

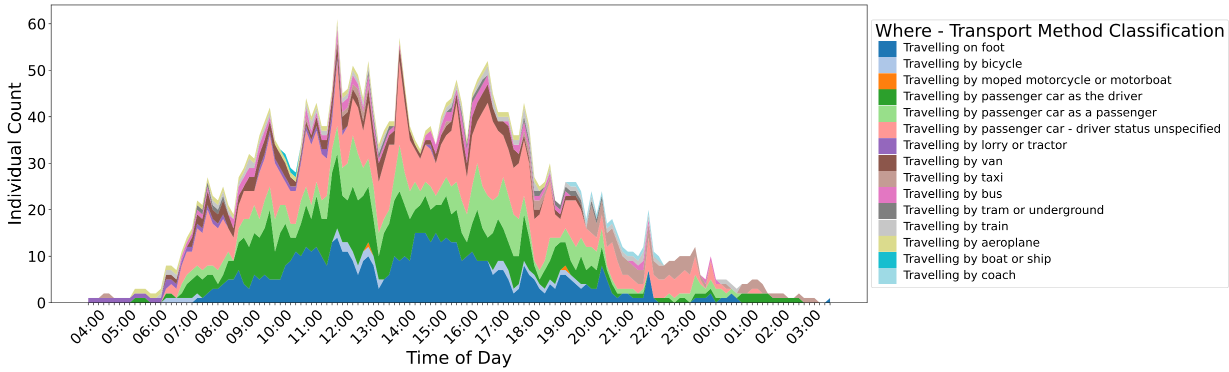
Supplementary Table S5: **Traffic flow average for road types for grid with centroid [52.218, -0.07] in South Cambridgeshire.** The overall grid score for each transport method for the grid is visualised in Figure S6. The grid score provides an overall indication of the traffic across all road types within the grid.



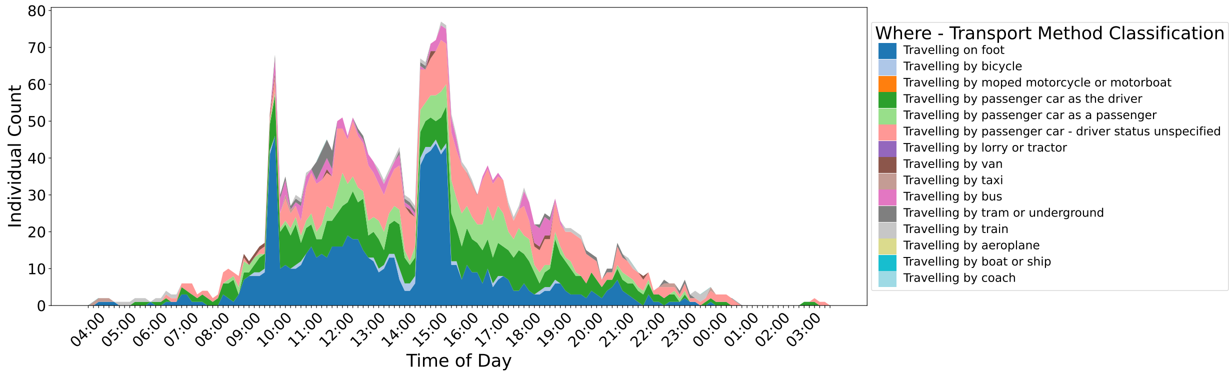
Supplementary Figure S6: **Road network infrastructure for a single grid with centroid [52.218, -0.07] in South Cambridgeshire.** The sub-road networks are shown for each of the five road types within the 1km^2 grid.



(a) Weekday

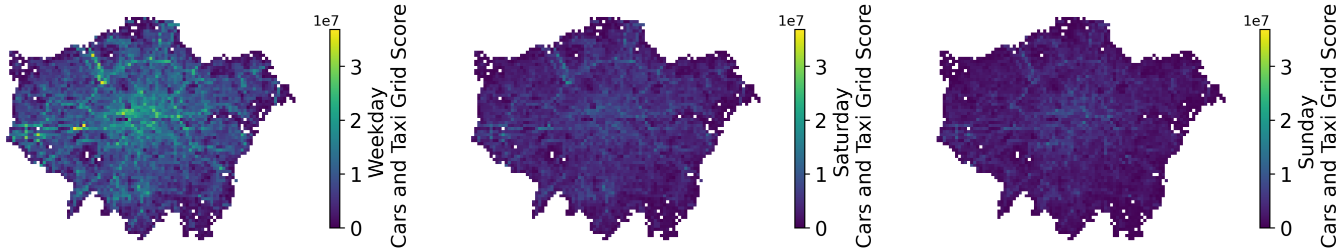


(b) Saturday

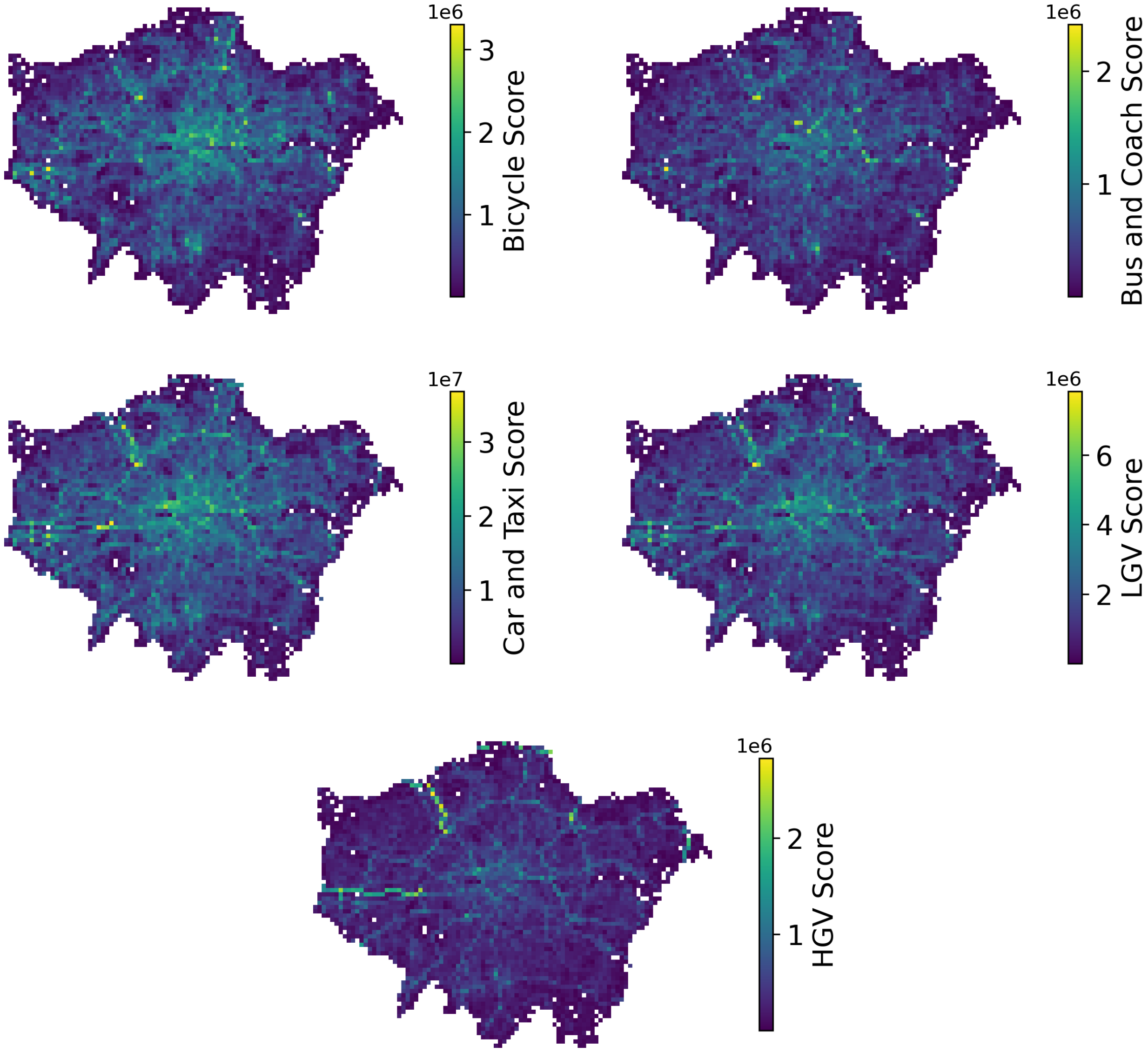


(c) Sunday

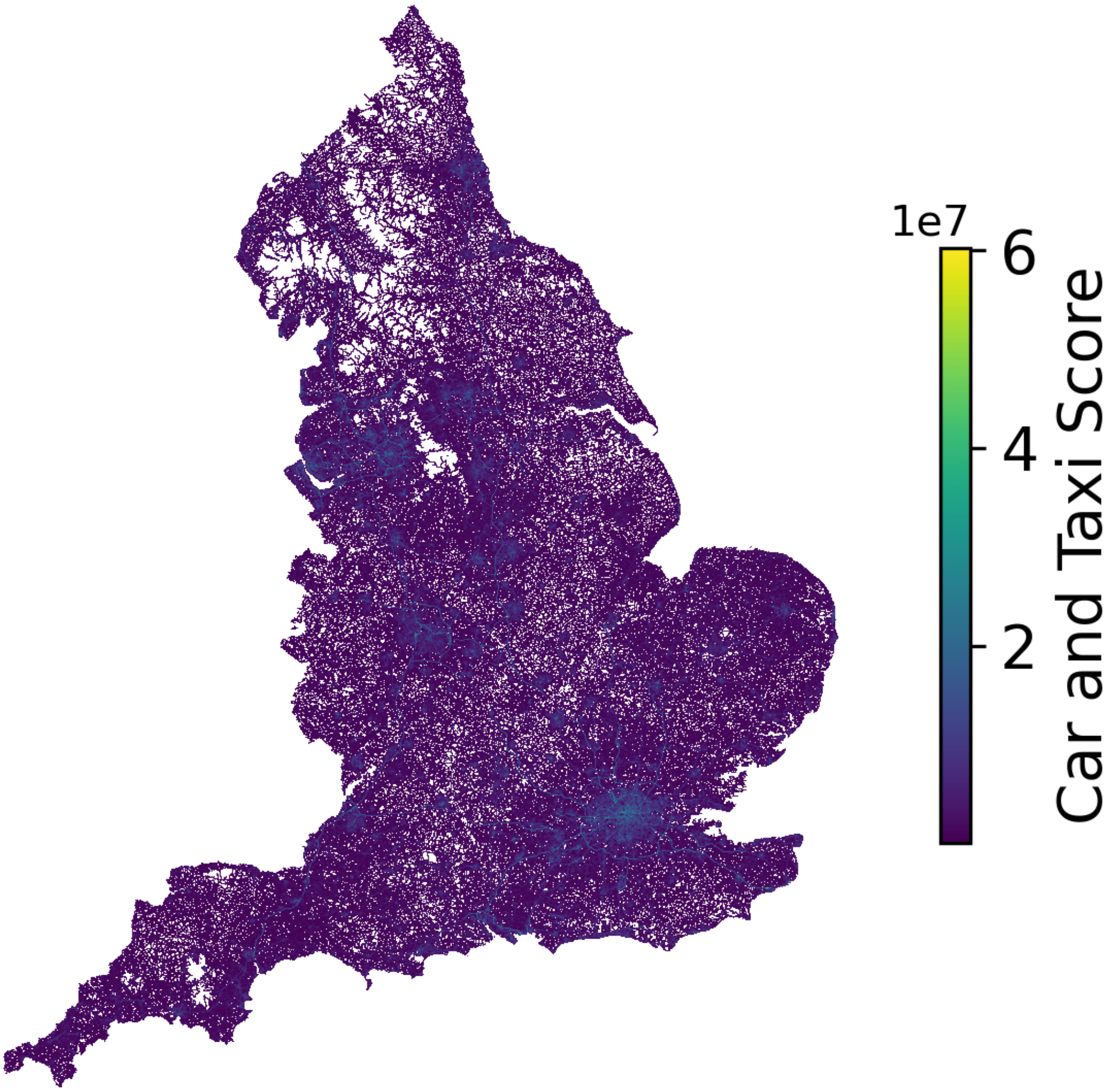
Supplementary Figure S7: **Spatial microsimulation for South Cambridgeshire 020 shows the proportion of different travel methods for all individuals travelling at the given time.** Of note are the different scales of the plots—total number of individuals taking a journey Weekday: 32622, Saturday: 16850, Sunday: 13857. Interestingly, even though Saturday has a higher overall number of individuals travelling, the peak on Sunday is higher.



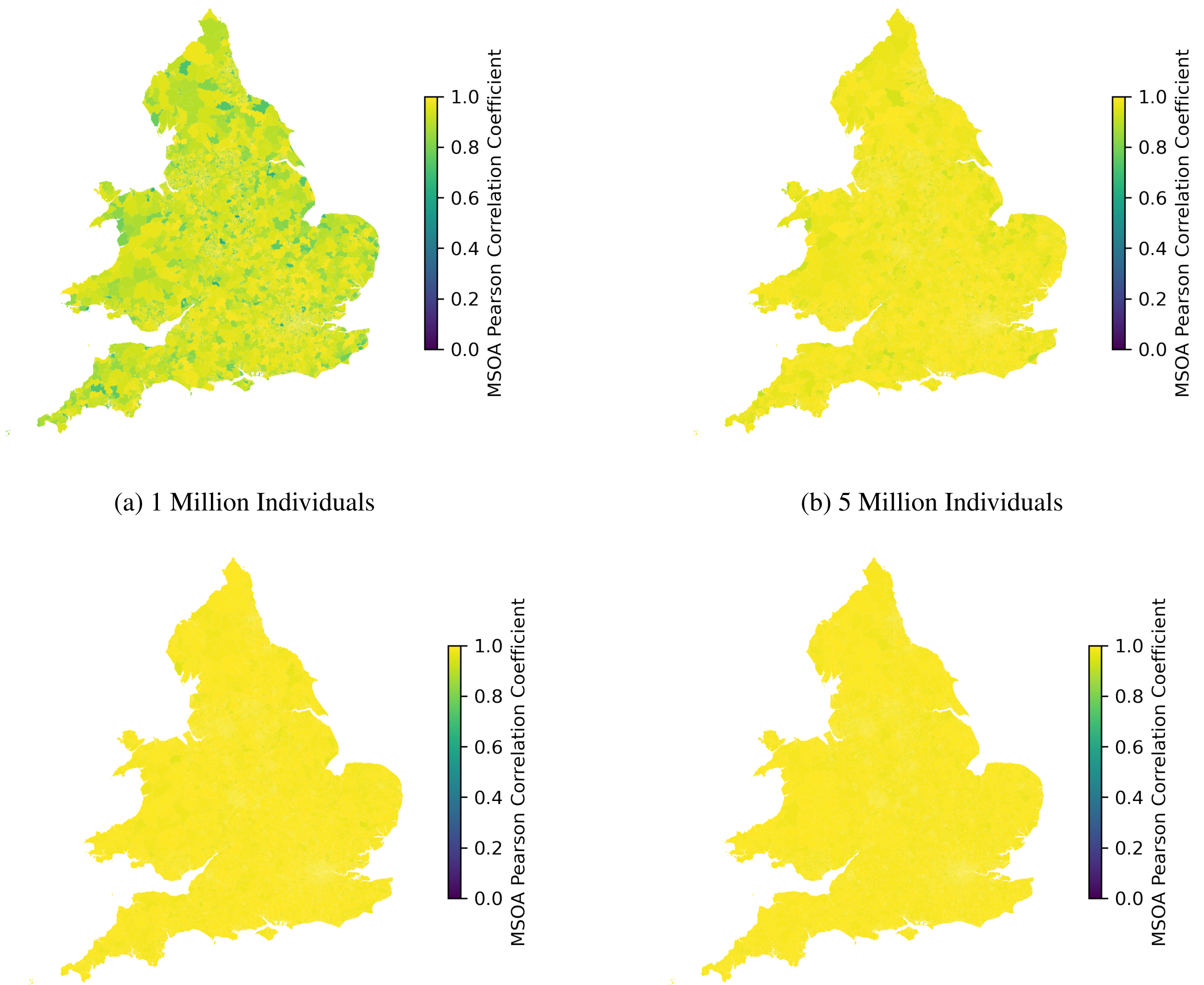
Supplementary Figure S8: **Temporal distribution of traffic grid score within london at 08AM on 1st, 2nd and 3rd June 2018 for cars and taxis.**



Supplementary Figure S9: **Transport use dataset for central London.** The transport use grids across central London help to highlight the difference in road usage across different road types across the five transport methods. Bicycle usage is more substantial in central London, with cars and taxis pervasive throughout and HGVs using the arterial main roads coming into the city.



Supplementary Figure S10: **Example complete England transport use dataset for Car and Taxi Score.** Of note is that not every grid is present within the figures, as not every 1km^2 grid within the study has any road infrastructure. The final feature vector for estimating the air pollution concentration is filled with zeros as those grids have no traffic.



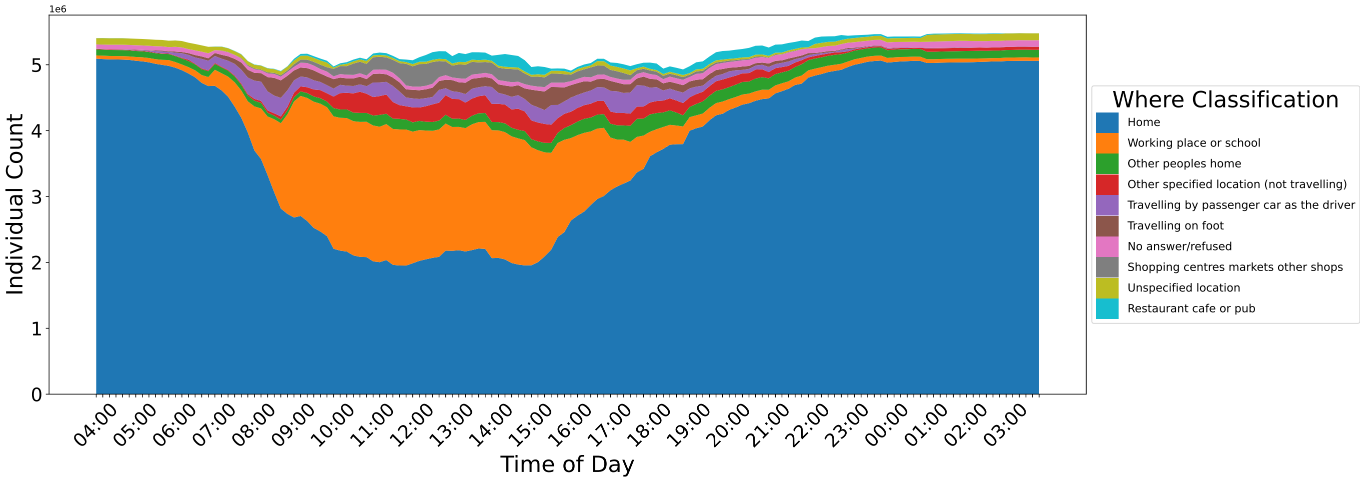
(a) 1 Million Individuals

(b) 5 Million Individuals

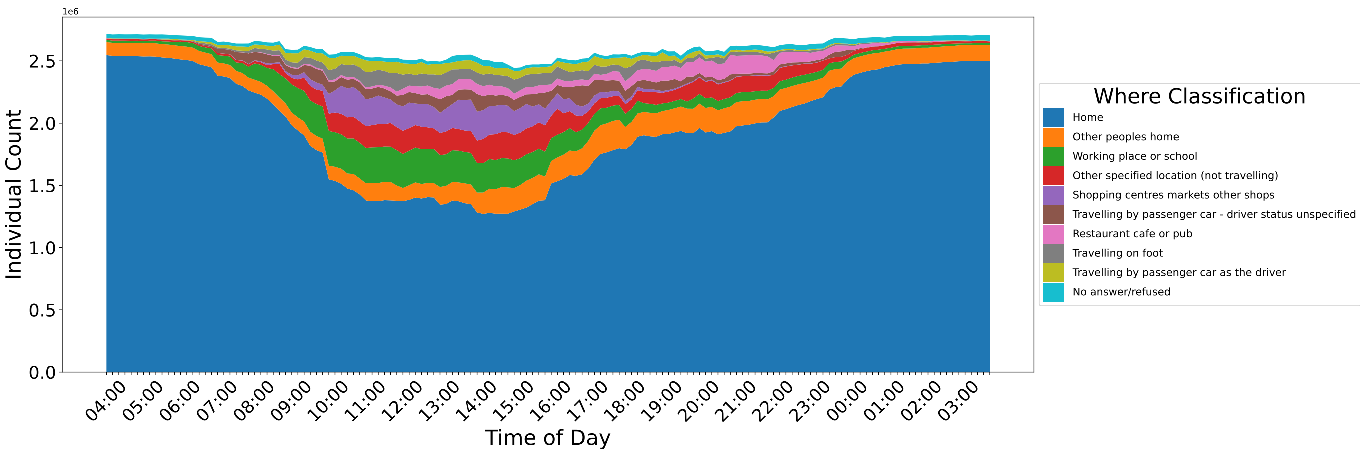
(c) 11 Million Individuals

(d) 55 Million Individuals

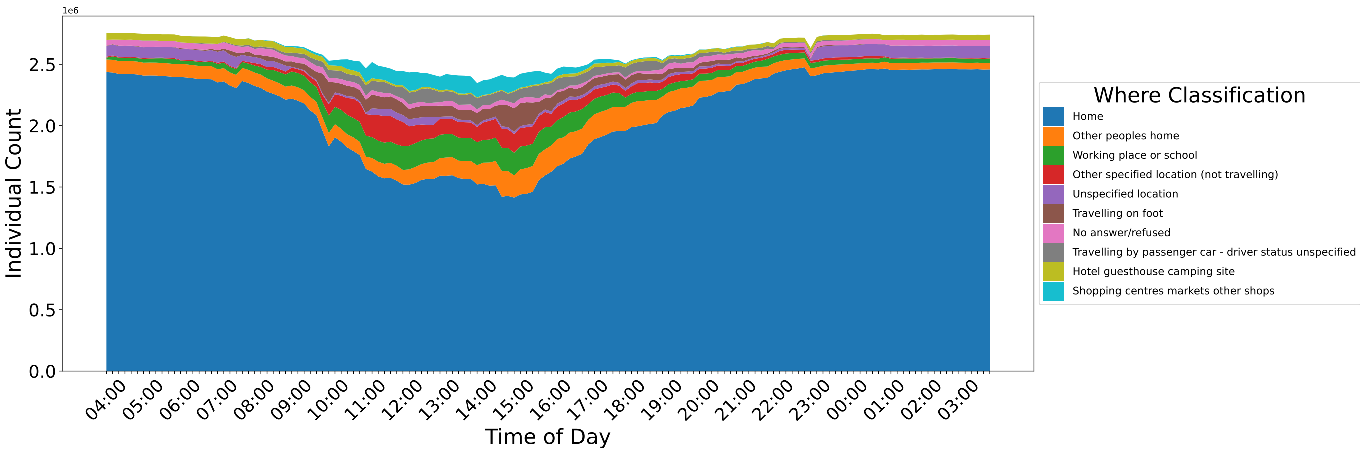
Supplementary Figure S11: **Pearson correlation coefficients across the MSOA regions used during the spatial microsimulation for different total numbers of simulated individuals.** During the spatial microsimulation, the number of individuals to be created was experimented with, intending to achieve a desirable Pearson Correlation coefficient across all MSOA regions ensuring model validity in population representation [9] while reducing the memory and computation burden associated with creating all individuals across the UK. While the simulation of all 55 million individuals in the UK resulted in a good pearson score, it was computationally expensive; as such, 11 million individuals were chosen to be simulated with the Pearson correlation being maintained at above 0.8 while significantly reducing computational costs.



(a) Weekday



(b) Saturday

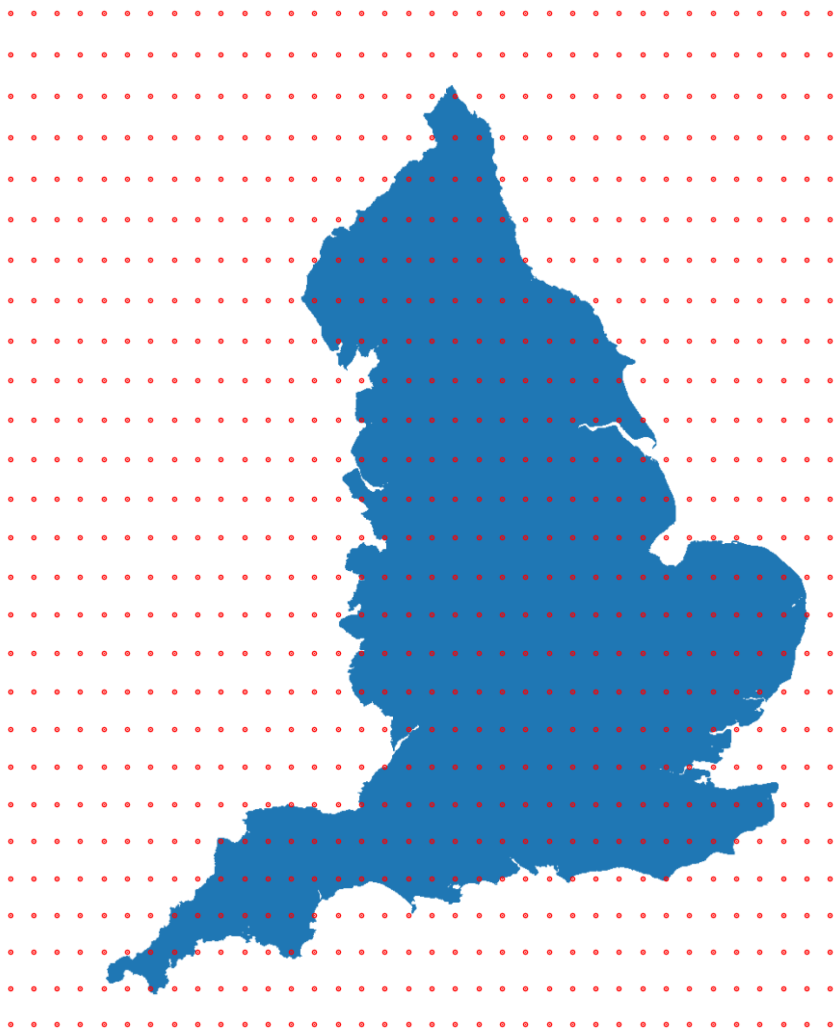


(c) Sunday

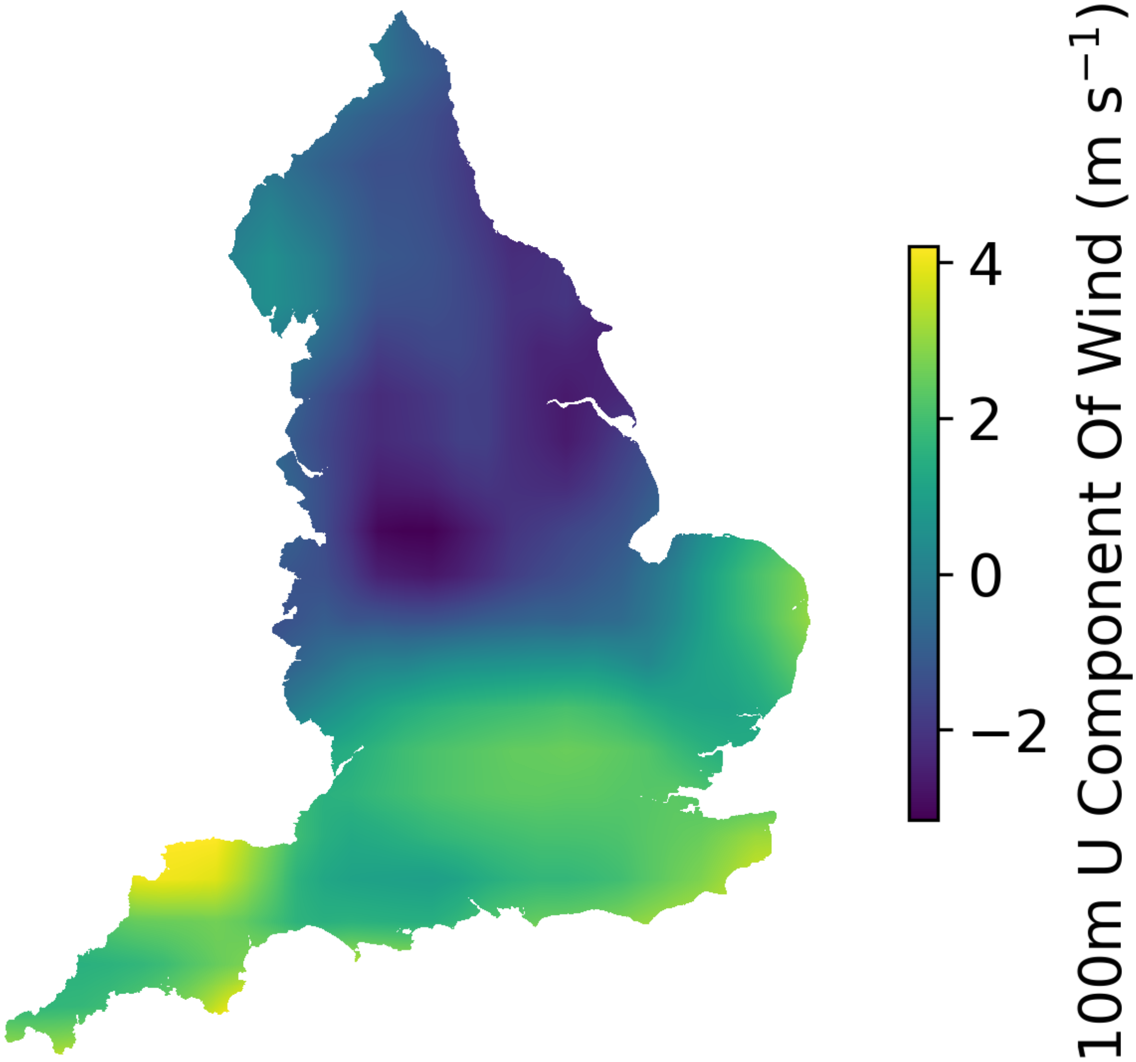
Supplementary Figure S12: **UK Time Use stack plots showing where individuals are across all the major location categories included in the dataset.**

S1.5 Meteorological

We retrieved meteorological data from the ECMWF Re-analysis version 5 (ERA5) dataset [10]. ERA5 is a global dataset that details the environmental conditions at a range of point locations worldwide. There are 100s of variables available through the data set; we chose a subset of 11 based on meteorological variables detailed as being strongly associated with air pollution concentration in the existing literature. The subset of 11 variables to include was the 100m and 10m U component of wind, the 100m and 10m V component of wind, 2m dewpoint temperature, 2m temperature, boundary layer height, downwards UV radiation at the surface, the instantaneous 10m wind gust, surface pressure and total column rainwater. To create the feature vector, the point locations within the ERA5 dataset seen in Figure S13 were interpolated across the study area to determine the variable value at each of the 1km^2 grid centroid. The resulting interpolated values at the grid centroids for a meteorological variable used, 100m U Component of Wind, are shown in Figure S14.



Supplementary Figure S13: **The blue region denotes the area of interest presented in Figure S1 with the red points showing the ERA5 sample locations across the UK.**



Supplementary Figure S14: **Example complete England meteorological dataset from ERA5 for the 100m U Component of Wind feature vector.**

S1.6 Remote Sensing

Google Earth Engine [11] derived datasets from Sentinel 5P [12] measurements comprised the remote sensing dataset family. The temporal period of datasets used was from 01-02-2019 to 01-03-2020, which allowed all available datasets from the Sentinel 5P platform to be studied. To ensure that all of the grids within the study area have a value for each timestamp, we aggregated the sentinel 5P datasets to the monthly mean temporal level. The grid would be interpolated from neighbouring values if any values were missing from the monthly aggregate, which was not the case for the variables used in the final study: NO_2, CO, HCHO, O_3, and the Absorbing Aerosol Index. Table S6 shows the number of missing data points across the study area for each month's different variables of consideration from Sentinel 5P. Methane (CH_4) was missing many data points and was therefore excluded from the dataset.

The process produced a spatially complete map of air pollution concentrations for each month of the year, which was then backfilled to other periods from before the Sentinel 5P platform came online to indicate typical air pollution concentrations during each month. Figure S15 shows the complete spatial map of the remote sensing dataset produced for June for NO_2.

Month	2019	2020	2021	2022	Monthly Median Overall	Monthly Median Overall Interpolated
1.00	-	100.00	99.66	100.00	100.00	100.00
2.00	100.00	100.00	100.00	-	100.00	100.00
3.00	100.00	100.00	100.00	-	100.00	100.00
4.00	100.00	100.00	100.00	-	100.00	100.00
5.00	100.00	100.00	100.00	-	100.00	100.00
6.00	100.00	100.00	100.00	-	100.00	100.00
7.00	100.00	99.92	100.00	-	100.00	100.00
8.00	100.00	100.00	100.00	-	100.00	100.00
9.00	100.00	100.00	100.00	-	100.00	100.00
10.00	100.00	99.87	100.00	-	100.00	100.00
11.00	100.00	100.00	100.00	-	100.00	100.00
12.00	100.00	99.98	100.00	-	100.00	100.00

(a) **NO_2**

Month	2019	2020	2021	2022	Monthly Median Overall	Monthly Median Overall Interpolated
1.00	-	100.00	73.88	96.09	100.00	100.00
2.00	100.00	100.00	100.00	-	100.00	100.00
3.00	100.00	100.00	100.00	-	100.00	100.00
4.00	100.00	100.00	100.00	-	100.00	100.00
5.00	100.00	100.00	100.00	-	100.00	100.00
6.00	100.00	100.00	100.00	-	100.00	100.00
7.00	100.00	100.00	100.00	-	100.00	100.00
8.00	100.00	100.00	100.00	-	100.00	100.00
9.00	100.00	100.00	100.00	-	100.00	100.00
10.00	100.00	99.98	100.00	-	100.00	100.00
11.00	100.00	100.00	100.00	-	100.00	100.00
12.00	99.95	0.99	0.25	-	99.99	100.00

(b) **HCHO**

Month	2019	2020	2021	2022	Monthly Median Overall	Monthly Median Overall Interpolated
1.00	-	100.00	100.00	100.00	100.00	100.00
2.00	100.00	100.00	100.00	-	100.00	100.00
3.00	100.00	100.00	100.00	-	100.00	100.00
4.00	100.00	100.00	100.00	-	100.00	100.00
5.00	100.00	100.00	100.00	-	100.00	100.00
6.00	100.00	100.00	100.00	-	100.00	100.00
7.00	100.00	100.00	100.00	-	100.00	100.00
8.00	100.00	100.00	100.00	-	100.00	100.00
9.00	100.00	100.00	100.00	-	100.00	100.00
10.00	100.00	100.00	100.00	-	100.00	100.00
11.00	100.00	100.00	100.00	-	100.00	100.00
12.00	100.00	100.00	100.00	-	100.00	100.00

(c) **O_3**

Month	2019	2020	2021	2022	Monthly Median Overall	Monthly Median Overall Interpolated
1.00	-	100.00	100.00	100.00	100.00	100.00
2.00	100.00	100.00	100.00	-	100.00	100.00
3.00	100.00	100.00	100.00	-	100.00	100.00
4.00	100.00	100.00	100.00	-	100.00	100.00
5.00	100.00	100.00	100.00	-	100.00	100.00
6.00	100.00	100.00	100.00	-	100.00	100.00
7.00	100.00	100.00	100.00	-	100.00	100.00
8.00	100.00	100.00	100.00	-	100.00	100.00
9.00	100.00	100.00	100.00	-	100.00	100.00
10.00	100.00	100.00	100.00	-	100.00	100.00
11.00	100.00	100.00	100.00	-	100.00	100.00
12.00	100.00	100.00	100.00	-	100.00	100.00

(d) **CO**

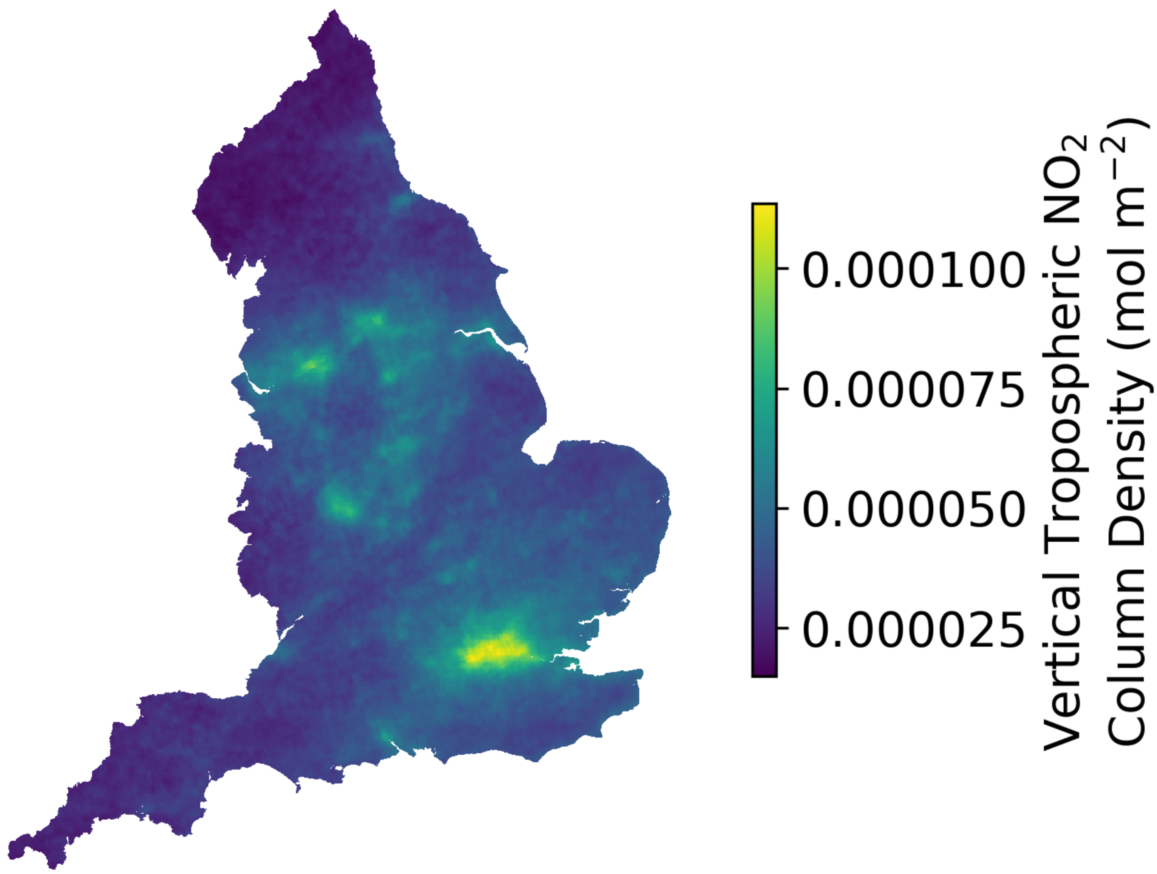
Month	2019	2020	2021	2022	Monthly Median Overall	Monthly Median Overall Interpolated
1.00	-	0.16	1.24	12.62	13.52	100.00
2.00	77.07	46.45	63.81	-	82.09	100.00
3.00	37.21	47.66	56.04	-	76.85	100.00
4.00	63.74	72.79	83.10	-	86.00	100.00
5.00	22.90	69.36	60.84	-	80.16	100.00
6.00	49.87	55.91	60.14	-	77.18	100.00
7.00	23.54	32.61	78.60	-	82.90	100.00
8.00	23.21	47.09	16.11	-	64.44	100.00
9.00	53.45	81.42	79.16	-	87.13	100.00
10.00	39.47	3.46	49.58	-	64.93	100.00
11.00	3.69	31.50	16.63	-	38.55	100.00

(e) **CH_4**

Month	2019	2020	2021	2022	Monthly Median Overall	Monthly Median Overall Interpolated
1.00	-	100.00	100.00	100.00	100.00	100.00
2.00	100.00	100.00	100.00	-	100.00	100.00
3.00	100.00	100.00	100.00	-	100.00	100.00
4.00	100.00	100.00	100.00	-	100.00	100.00
5.00	100.00	100.00	100.00	-	100.00	100.00
6.00	100.00	100.00	100.00	-	100.00	100.00
7.00	100.00	100.00	100.00	-	100.00	100.00
8.00	100.00	100.00	100.00	-	100.00	100.00
9.00	100.00	100.00	100.00	-	100.00	100.00
10.00	100.00	100.00	100.00	-	100.00	100.00
11.00	100.00	100.00	100.00	-	100.00	100.00
12.00	100.00	100.00	100.00	-	100.00	100.00

(f) **Absorbing Aerosol Index**

Supplementary Table S6: **Missing Data for each of the months for each of the variables considered in the study.** All of the variables considered other than CH_4 had a spatially complete dataset at the monthly aggregate level and, as such, were included in the study as shown in Figure S15.



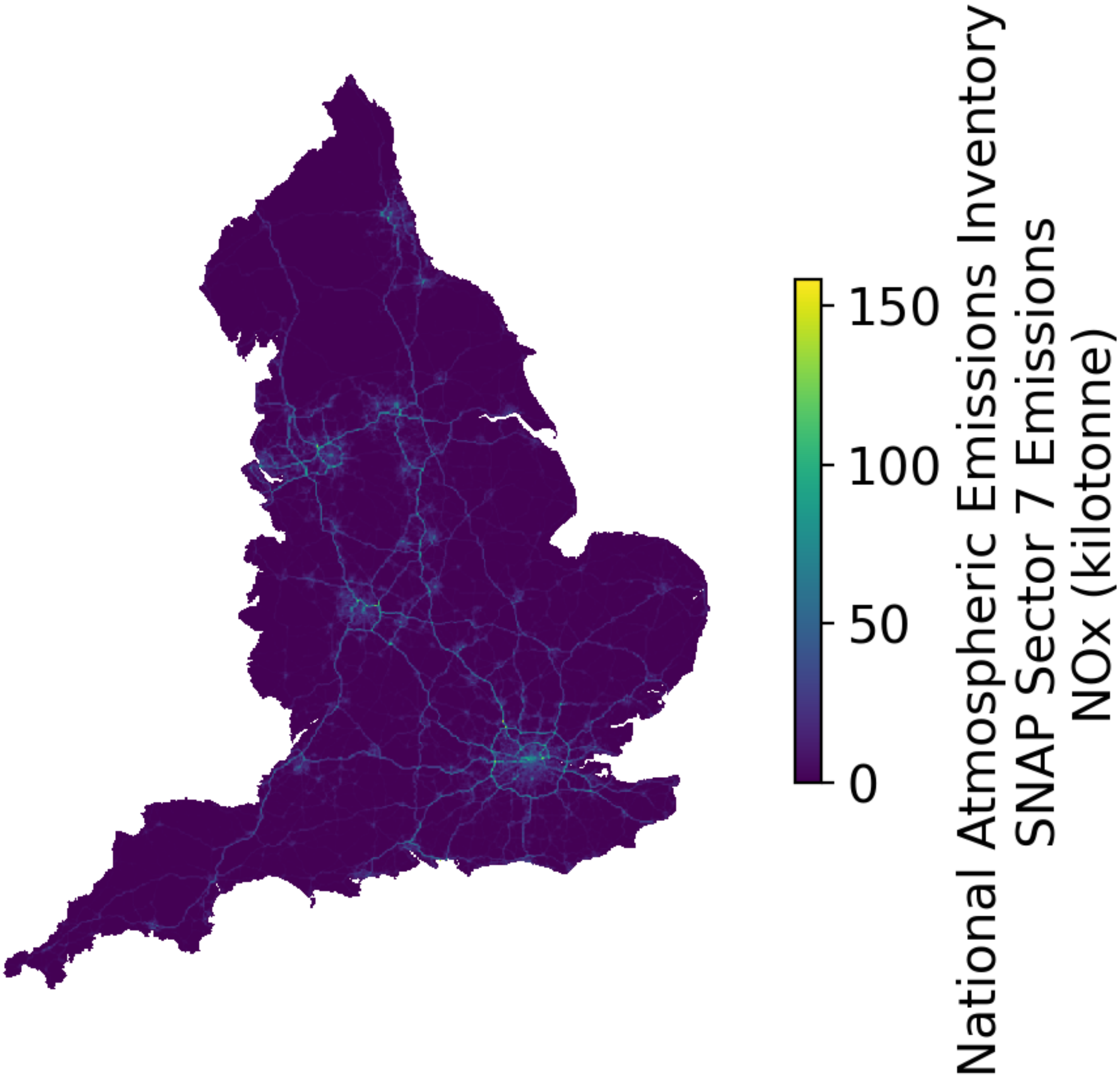
Supplementary Figure S15: **Example complete England remote sensing dataset from Sentinel 5P Google Earth Engine for NO_2.**

S1.7 Emissions

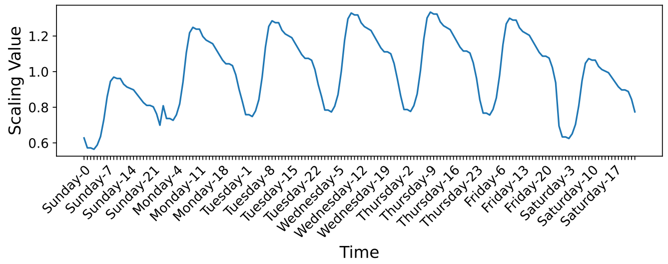
Emissions data is gathered from the UK National Atmospheric Emissions Inventory (NAEI) [13]. A set of seven air pollutants are included: $PM_{2.5}$, PM_{10}, (Non-methane volatile organic compounds (NMVOC), NH_3, SO_x, CO, NO_x in the study. The emissions are classified into one of 11 sectors to denote the emission source, based on Selected Nomenclature for Air Pollutants (SNAP) sectors [14]:

- SNAP Sector 1 (Combustion Energy Production and Transformation)
- SNAP Sector 2 (Combustion in Commercial, Institutional, Residential and Agriculture)
- SNAP Sector 3 (Combustion in Industry)
- SNAP Sector 4 (Production Processes)
- SNAP Sector 5 (Extraction and Distribution of Fossil Fuels)
- SNAP Sector 6 (Solvent Use)
- SNAP Sector 7 (Road Transport)
- SNAP Sector 8 (Other Transport and Mobile Machinery)
- SNAP Sector 9 (Waste Treatment and Disposal)
- SNAP Sector 10 (Agriculture, Forestry and Land Use Change)
- SNAP Sector 11 (Nature)

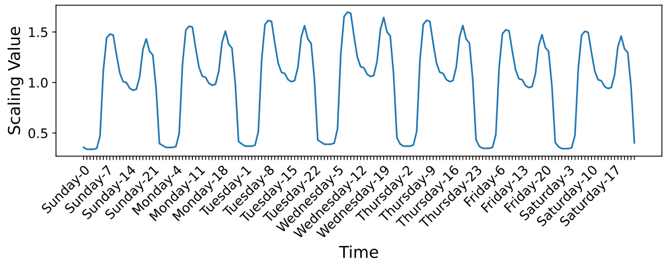
We created the emissions feature vector by summing the point and area emissions data [15] from the NAEI for each year per SNAP sector per species to map the emissions across the study area. We then subsequently scaled the emissions map depending on the timestamp of interest, applying a scaling for the hour and day of the week and month of interest [16]. Figure S16 gives an example of the emission feature vector for NO_x SNAP Sector 7 (Road) emissions across the study area.



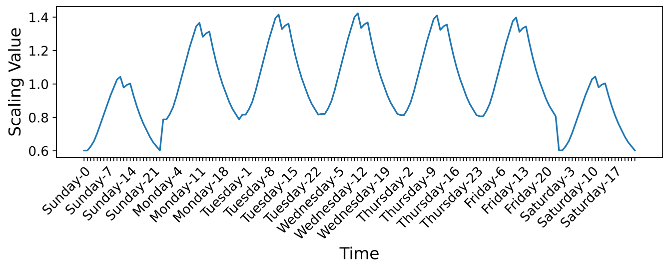
Supplementary Figure S16: **Example complete England emission dataset for SNAP Sector 7 (Road Emissions) for NO_x on 1st June 2018 at 0800AM.**



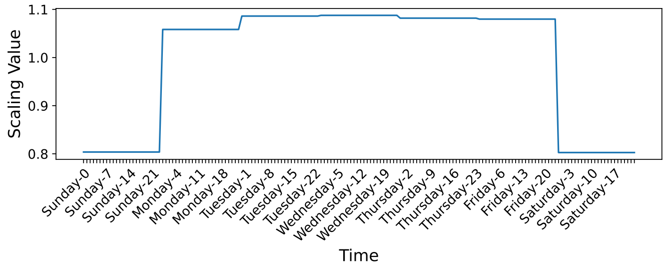
(a) SNAP Sector 1



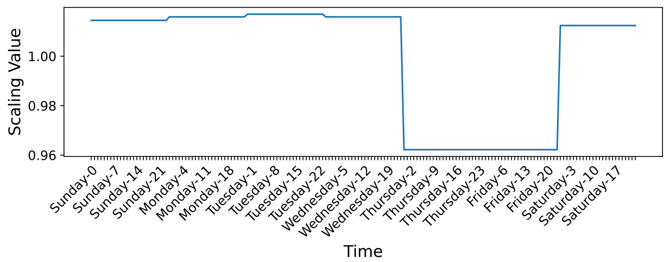
(b) SNAP Sector 2



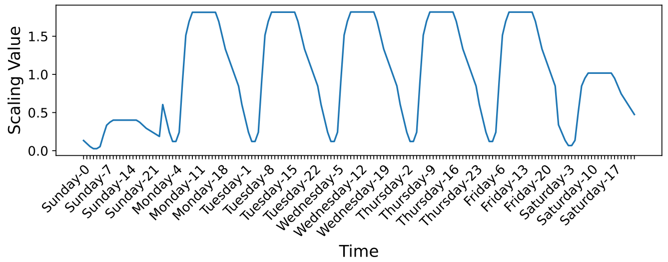
(c) SNAP Sector 3



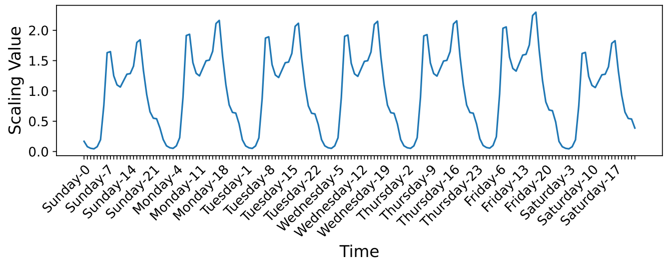
(d) SNAP Sector 4



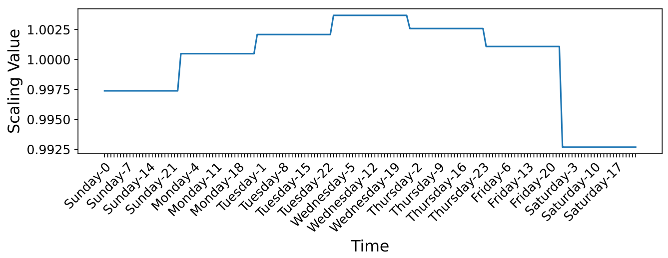
(e) SNAP Sector 5



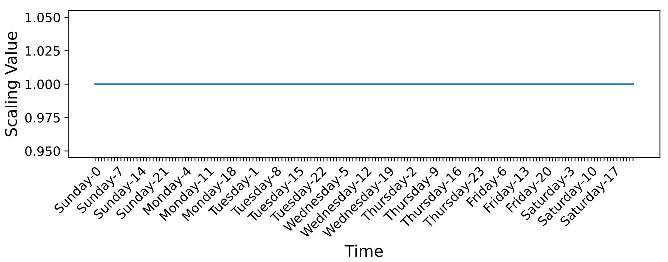
(f) SNAP Sector 6



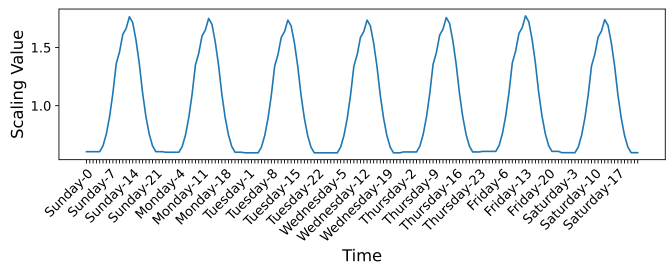
(g) SNAP Sector 7



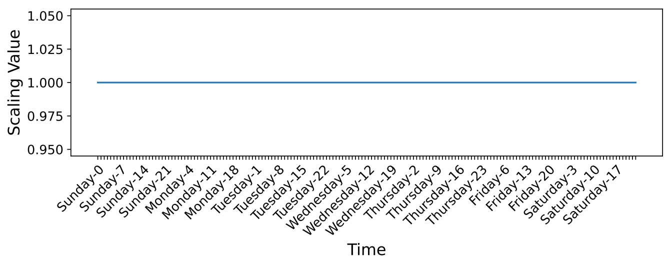
(h) SNAP Sector 8



(i) SNAP Sector 9

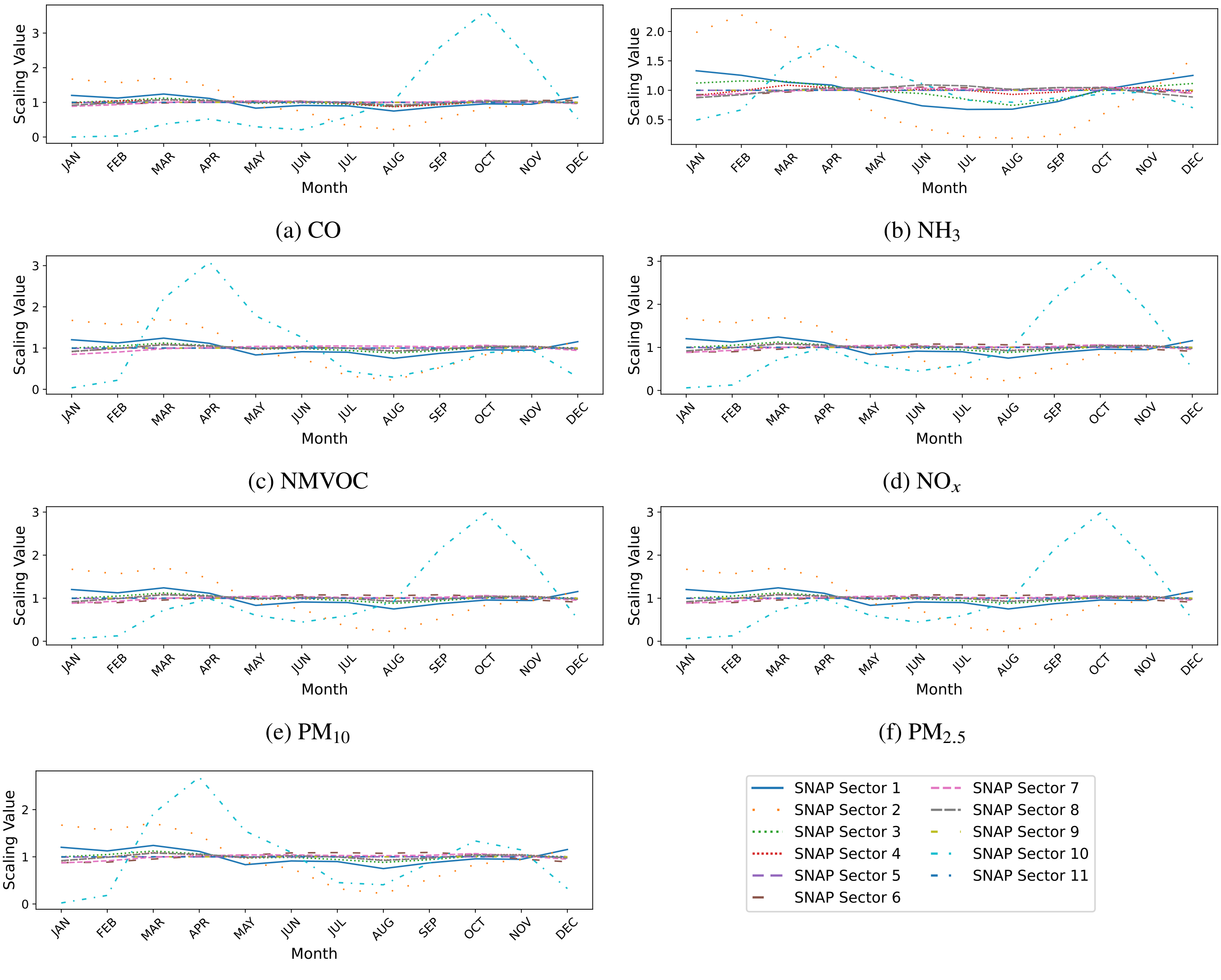


(j) SNAP Sector 10



(k) SNAP Sector 11

Supplementary Figure S17: **Hourly and daily emissions scaling per SNAP sector.**



(a) CO

(b) NH_3

(c) NMVOC

(d) NO_x

(e) PM_{10}

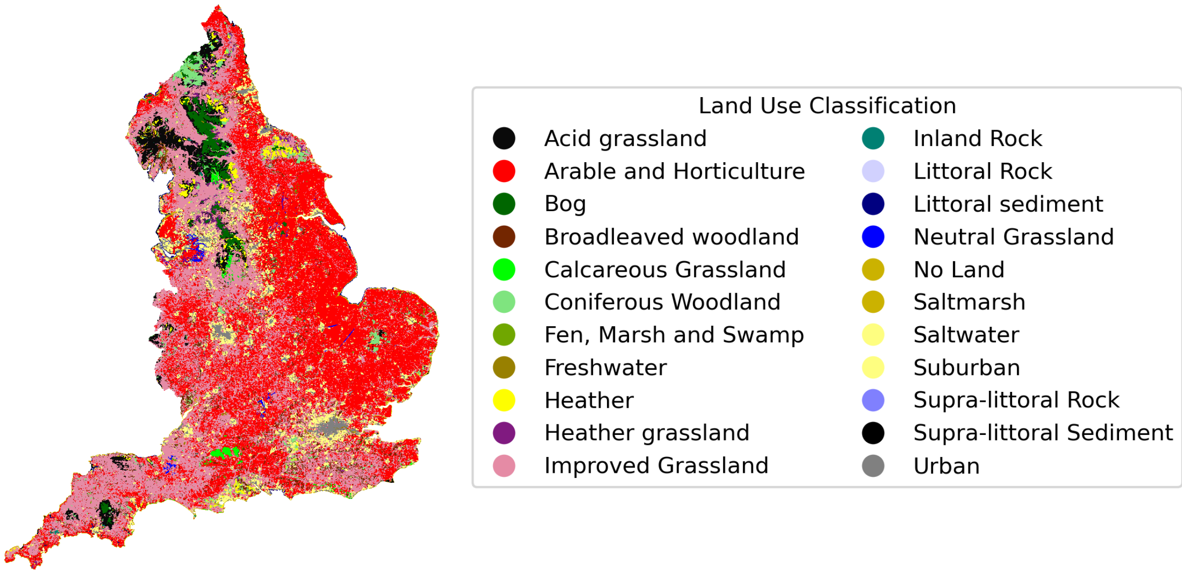
(f) $PM_{2.5}$

(g) SO_x

Supplementary Figure S18: **Emissions scaling for each emissions species by SNAP sector across the months.**

S1.8 Land Use

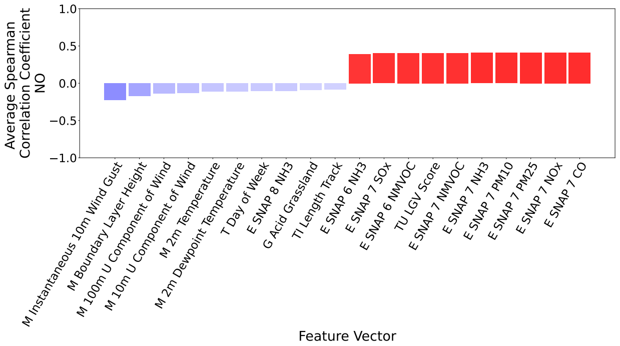
We created a geographic profile based on land use for each grid within the study. The 25m UKCEH Land Cover Maps [17] were used to profile each grid's land use composition across 22 possible land use classifications. The feature vector elements represent the number of pixels with a given land cover classification in the raster. Figure S19 shows the majority classification for each grid within the study.



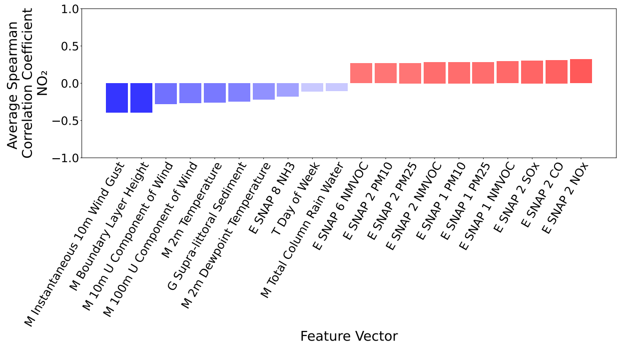
Supplementary Figure S19: **Land use majority classification per grid.**

S2 Feature Selection

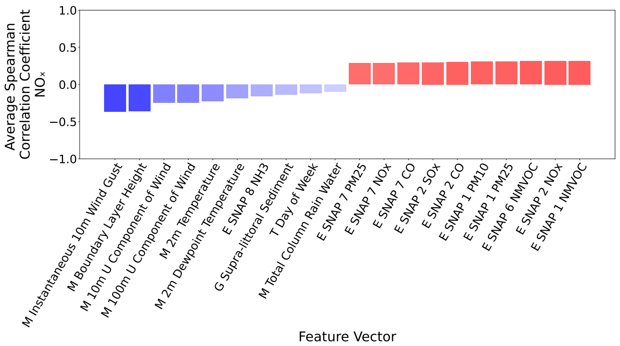
S2.1 Air Pollutants and Feature Vector



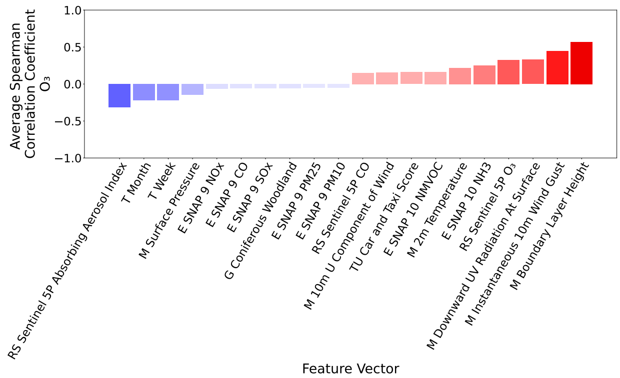
(a) NO



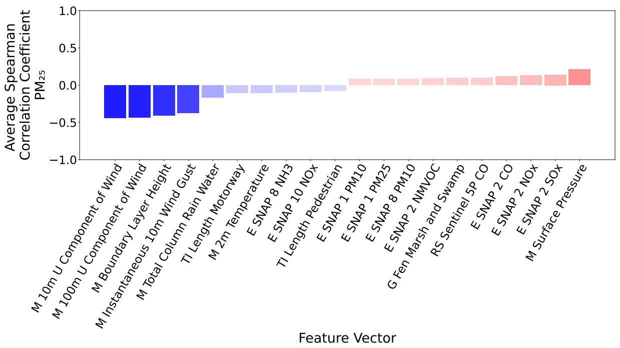
(b) NO_2



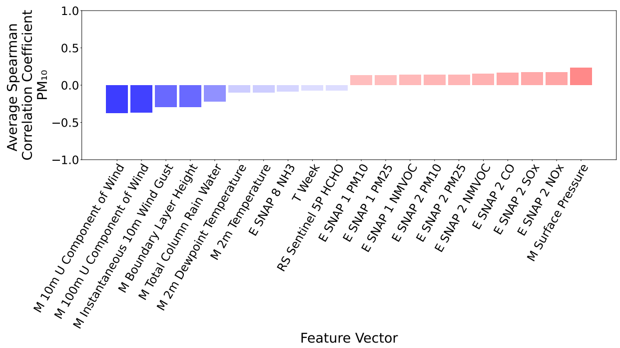
(c) NO_x



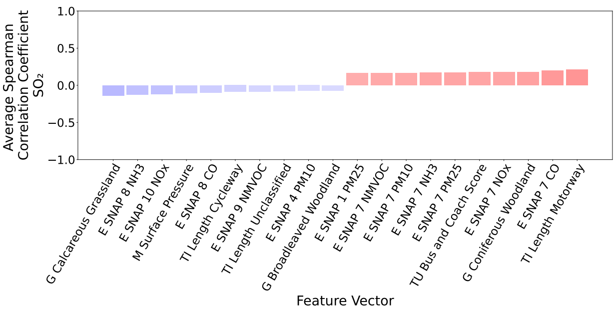
(d) O_3



(e) $PM_{2.5}$



(f) PM_{10}



(g) SO_2

Supplementary Figure S20: **Spearman correlation coefficients overall mean for all pollutants.** Of note is that the remote sensing dataset family does not have the highest correlation with the air pollutant that it measures directly, such as NO_2. The reason for this is the difference in temporal period, with the dataset being a monthly aggregate rather than the higher temporal resolution datasets included, such as the emissions datasets.

S2.2 Inter Feature Vectors

Linkage Distance	Number of Clusters
-1.00	139.00
0.05	112.00
0.10	103.00
0.25	83.00
0.50	62.00
0.75	51.00
1.00	29.00
1.25	15.00
1.50	11.00
1.75	8.00
2.00	6.00
2.50	4.00
3.00	1.00

Supplementary Table S7: **Linkage distance for hierarchial clustering of the feature vectors.** The linkage distance column provides a set of thresholds and the associated number of clusters grouped when all features under the threshold are put into a cluster. The smaller the linkage, the more similar the feature vectors are—the -1.00 threshold indicates that all feature vectors should be individual clusters, resulting in 139 clusters, with 4 missing due to land use and 9 emissions feature vectors not being present in a location with a monitoring station; providing the 152 feature vectors elements considered in the study.

S3 Modelling

S3.1 Model Performance Summary Analysis for Individual Monitoring Stations

Monitoring Station	Correlation	Bias ($\mu g/m^3$)	MSE $(\mu g/m^3)^2$
Aston Hill	0.80	-0.50	8.54
Barnsley Gawber	0.90	-0.74	36.10
Billingham	0.85	-1.42	69.63
Birkenhead Borough Road	0.89	-1.26	72.34
Birmingham A4540 Roadside	0.90	-0.91	65.07
Birmingham Acocks Green	0.92	-0.91	65.98
Blackburn Accrington Road	0.89	-0.96	52.71
Blackpool Marton	0.87	-1.00	41.71
Bournemouth	0.88	-0.98	34.83
Bradford Mayo Avenue	0.91	-1.22	130.54
Brighton Preston Park	0.85	-1.40	59.80
Bristol St Paul's	0.88	-1.24	75.79
Bury Whitefield Roadside	0.89	-1.22	74.80
Cambridge Roadside	0.89	-0.67	49.68
Camden Kerbside	0.90	-1.41	207.09
Cannock A5190 Roadside	0.89	-1.17	67.47
Canterbury	0.86	-0.96	32.88
Carlisle Roadside	0.86	-1.31	74.25
Charlton Mackrell	0.82	-0.64	12.22
Chatham Roadside	0.88	-1.12	59.96
Chesterfield Loundsley Green	0.87	-0.87	31.27
Chesterfield Roadside	0.88	-0.85	42.91
Chilbolton Observatory	0.89	-0.58	19.90
Christchurch Barrack Road	0.86	-2.01	118.80
Coventry Allesley	0.90	-0.98	54.34
Doncaster A630 Cleveland Street	0.88	-0.96	60.90
Eastbourne	0.87	-0.95	32.34
Exeter Roadside	0.88	-1.41	92.45
Glazebury	0.84	-1.40	50.93
Haringey Roadside	0.90	-0.68	73.71
High Muffles	0.77	-0.83	17.02
Honiton	0.84	-0.73	18.65
Horley	0.85	-1.34	57.55
Hull Freetown	0.88	-1.03	53.35
Hull Holderness Road	0.88	-1.24	83.57
Ladybower	0.78	-0.86	18.35
Leamington Spa	0.90	-0.79	39.55
Leamington Spa Rugby Road	0.92	-0.73	35.42
Leeds Centre	0.88	-1.07	71.68
Leeds Headingley Kerbside	0.91	-1.26	102.35
Leicester A594 Roadside	0.90	-1.19	99.00
Leicester University	0.88	-1.10	62.12
Leominster	0.85	-0.72	18.30
Liverpool Speke	0.88	-1.09	58.38
London Bexley	0.90	-1.01	58.44
London Bloomsbury	0.89	-0.97	88.76
London Eltham	0.90	-0.78	38.55
London Haringey Priory Park South	0.91	-0.86	47.08
London Harlington	0.89	-1.42	92.07
London Hillingdon	0.89	-1.52	201.58
London Marylebone Road	0.89	-1.16	336.94

(a) **Mean, Max and Minimum values for Bias, Correlation and MSE for NO_2 across all air pollution monitoring stations.**

S3.2 Data Subsetting - Temporal

Monitoring Station	Correlation	Bias (μg/m^3)	MSE (μg/m^3)^2
London N. Kensington	0.91	-0.97	73.71
London Westminster	0.91	-1.01	80.54
Lullington Heath	0.75	-1.05	23.48
Luton A505 Roadside	0.90	-1.51	176.31
Manchester Piccadilly	0.87	-0.95	83.31
Manchester Sharston	0.87	-0.99	58.94
Middlesbrough	0.87	-1.13	47.40
Newcastle Centre	0.88	-0.79	56.64
Newcastle Cradlewell Roadside	0.87	-1.97	192.65
Norwich Lakenfields	0.87	-0.73	24.30
Nottingham Centre	0.90	-0.75	53.21
Nottingham Western Boulevard	0.88	-1.44	113.77
Oldbury Birmingham Road	0.84	-1.88	137.28
Oxford Centre Roadside	0.90	-1.13	134.48
Oxford St Ebbes	0.89	-0.84	32.05
Plymouth Centre	0.84	-1.47	75.55
Preston	0.89	-0.99	54.58
Reading New Town	0.90	-0.86	61.13
Rochester Stoke	0.86	-1.10	38.07
Salford Eccles	0.89	-1.12	67.86
Sandy Roadside	0.87	-1.34	102.95
Scunthorpe Town	0.88	-1.16	59.37
Shaw Crompton Way	0.88	-1.24	91.71
Sheffield Barnsley Road	0.89	-0.99	115.26
Sheffield Devonshire Green	0.88	-1.11	68.71
Sheffield Tinsley	0.88	-1.23	76.07
Southampton A33	0.85	-2.48	196.13
Southampton Centre	0.83	-1.32	87.31
Southend-on-Sea	0.88	-0.98	46.74
Southwark A2 Old Kent Road	0.89	-1.62	180.87
St Helens Linkway	0.87	-1.49	112.89
St Osyth	0.82	-1.01	33.72
Stanford-le-Hope Roadside	0.90	-1.11	67.22
Stockton-on-Tees A1305 Roadside	0.87	-1.38	69.30
Stockton-on-Tees Eaglescliffe	0.85	-1.47	64.43
Stoke-on-Trent A50 Roadside	0.88	-1.88	243.40
Stoke-on-Trent Centre	0.89	-1.02	55.38
Storrington Roadside	0.87	-0.91	57.98
Sunderland Silksworth	0.87	-1.08	39.29
Sunderland Wessington Way	0.88	-1.50	75.43
Thurrock	0.88	-1.21	67.33
Tower Hamlets Roadside	0.89	-1.23	144.76
Walsall Woodlands	0.90	-0.88	43.30
Warrington	0.83	-1.27	62.00
Wicken Fen	0.85	-0.71	16.43
Widnes Milton Road	0.88	-1.88	154.93
Wigan Centre	0.90	-0.97	45.24
Wirral Tranmere	0.87	-1.20	56.98
Worthing A27 Roadside	0.86	-0.98	114.12
Yarner Wood	0.77	-0.51	7.65
York Bootham	0.87	-1.10	38.49
York Fishergate	0.88	-0.92	56.24

(b) **Mean, Max and Minimum values for Bias, Correlation and MSE for NO_2 across all air pollution monitoring stations. (cont.)**

Supplementary Table S8

Monitoring Station	Correlation	Bias (μg/m^3)	MSE (μg/m^3)^2
Aston Hill	0.82	-0.87	98.44
Barnsley Gawber	0.89	-1.46	87.07
Birmingham A4540 Roadside	0.91	-1.95	92.65
Birmingham Acocks Green	0.91	-1.28	91.60
Blackpool Marton	0.88	-1.86	124.78
Bournemouth	0.89	-2.08	141.40
Brighton Preston Park	0.87	-2.62	164.38
Bristol St Paul's	0.89	-1.90	126.46
Canterbury	0.89	-2.35	155.53
Charlton Mackrell	0.88	-0.70	106.57
Chilbolton Observatory	0.91	-1.64	107.98
Coventry Allesley	0.91	-1.77	99.18
Exeter Roadside	0.86	-1.86	92.98
Glazebury	0.91	-1.97	108.71
High Muffles	0.84	-1.37	121.79
Hull Freetown	0.88	-1.77	121.74
Ladybower	0.86	-1.25	95.41
Leamington Spa	0.91	-1.49	101.01
Leeds Centre	0.88	-1.94	119.62
Leicester University	0.91	-1.70	106.17
Leominster	0.90	-1.62	118.16
Liverpool Speke	0.89	-1.63	107.55
London Bloomsbury	0.88	-2.12	100.38
London Eltham	0.91	-1.75	103.00
London Haringey Priory Park South	0.91	-1.89	112.01
London Harlington	0.91	-2.12	114.93
London Hillingdon	0.90	-2.28	106.30
London Marylebone Road	0.88	-1.32	47.30
London N. Kensington	0.91	-2.04	115.81
Lullington Heath	0.84	-1.61	123.50
Manchester Piccadilly	0.88	-1.82	86.26
Manchester Sharston	0.89	-2.19	123.28
Middlesbrough	0.86	-1.90	117.96
Newcastle Centre	0.87	-1.84	116.64
Norwich Lakenfields	0.89	-1.43	100.70
Nottingham Centre	0.90	-1.60	92.25
Plymouth Centre	0.85	-2.29	153.77
Preston	0.89	-1.62	110.04
Reading New Town	0.92	-1.77	95.97
Rochester Stoke	0.90	-1.66	117.77
Sheffield Devonshire Green	0.89	-1.70	108.01
Sibton	0.87	-1.47	112.02
Southampton Centre	0.87	-2.35	134.80
Southend-on-Sea	0.89	-1.38	113.44
St Osyth	0.88	-1.70	117.05
Stoke-on-Trent Centre	0.88	-1.60	104.48
Sunderland Silksworth	0.85	-1.82	133.16
Thurrock	0.90	-2.30	126.56
Walsall Woodlands	0.91	-1.48	105.39
Weybourne	0.84	-1.42	129.64
Wicken Fen	0.91	-1.33	98.00
Wigan Centre	0.90	-2.35	134.85
Wirral Tranmere	0.89	-1.55	105.35
Yarner Wood	0.85	-1.09	118.42

Supplementary Table S9: **Mean, Max and Minimum values for Bias, Correlation and MSE for O_3 across all air pollution monitoring stations.**

Monitoring Station	Correlation	Bias (µg/m^3)	MSE (µg/m^3)^2
Aston Hill	0.79	-0.53	11.02
Barnsley Gawber	0.84	-2.07	370.62
Billingham	0.79	-3.45	546.38
Birkenhead Borough Road	0.85	-3.22	615.12
Birmingham A4540 Roadside	0.86	-5.30	1600.53
Birmingham Acocks Green	0.88	-2.01	371.06
Blackburn Accrington Road	0.84	-3.73	809.36
Blackpool Marton	0.79	-1.91	199.60
Bournemouth	0.82	-2.00	234.64
Bradford Mayo Avenue	0.88	-5.66	2328.79
Brighton Preston Park	0.75	-4.02	729.92
Bristol St Paul’s	0.83	-4.92	1472.84
Bury Whitefield Roadside	0.85	-4.20	982.29
Cambridge Roadside	0.89	-4.80	1149.87
Camden Kerbside	0.89	-8.08	3855.19
Cannock A5190 Roadside	0.86	-2.78	454.75
Canterbury	0.83	-2.00	204.80
Carlisle Roadside	0.84	-5.71	1442.72
Charlton Mackrell	0.80	-0.82	26.69
Chatham Roadside	0.86	-2.93	470.84
Chesterfield Loundsley Green	0.84	-2.51	313.16
Chesterfield Roadside	0.85	-2.95	528.86
Chilbolton Observatory	0.83	-1.21	97.55
Christchurch Barrack Road	0.80	-6.78	1525.95
Coventry Allesley	0.88	-2.92	466.95
Doncaster A630 Cleveland Street	0.84	-3.12	703.57
Eastbourne	0.78	-2.10	214.99
Exeter Roadside	0.86	-6.46	1708.95
Glazebury	0.81	-3.13	406.70
Haringey Roadside	0.88	-4.04	1443.21
High Muffles	0.75	-0.88	25.04
Honiton	0.79	-1.09	69.70
Horley	0.84	-3.53	592.55
Hull Freetown	0.86	-2.17	331.12
Hull Holderness Road	0.86	-4.04	949.09
Ladybower	0.79	-0.95	28.32
Leamington Spa	0.81	-1.86	344.30
Leamington Spa Rugby Road	0.88	-1.90	326.60
Leeds Centre	0.83	-3.33	826.58
Leeds Headingley Kerbside	0.89	-4.62	1346.05
Leicester A594 Roadside	0.89	-4.49	1251.10
Leicester University	0.87	-2.94	480.34
Leominster	0.78	-1.35	110.83
Liverpool Speke	0.81	-2.81	426.80
London Bexley	0.87	-3.46	796.42
London Bloomsbury	0.86	-3.25	858.09
London Eltham	0.88	-2.16	362.21
London Haringey Priory Park South	0.88	-2.43	542.65
London Harlington	0.88	-4.88	1360.56
London Hillingdon	0.88	-6.22	2404.50
London Marylebone Road	0.90	-8.37	7587.10

(a) **Mean, Max and Minimum values for Bias, Correlation and MSE for NO_X across all air pollution monitoring stations.**

Monitoring Station	Correlation	Bias ($\mu g/m^3$)	MSE ($(\mu g/m^3)^2$)
London N. Kensington	0.88	-3.64	1057.42
London Westminster	0.89	-2.58	601.73
Lullington Heath	0.72	-1.14	33.97
Luton A505 Roadside	0.89	-7.03	3246.37
Manchester Piccadilly	0.83	-4.11	1133.80
Manchester Sharston	0.78	-4.34	1115.62
Middlesbrough	0.78	-2.70	423.03
Newcastle Centre	0.82	-2.51	625.20
Newcastle Cradlewell Roadside	0.85	-9.22	3033.26
Norwich Lakenfields	0.84	-1.34	129.67
Nottingham Centre	0.83	-2.96	791.20
Nottingham Western Boulevard	0.87	-6.29	1935.04
Oldbury Birmingham Road	0.81	-6.99	2294.37
Oxford Centre Roadside	0.88	-7.31	3122.44
Oxford St Ebbes	0.84	-3.52	574.63
Plymouth Centre	0.77	-4.48	797.49
Preston	0.84	-2.17	330.38
Reading New Town	0.87	-2.90	688.59
Rochester Stoke	0.81	-1.92	180.32
Salford Eccles	0.83	-4.25	1164.24
Sandy Roadside	0.84	-4.40	1200.23
Scunthorpe Town	0.84	-2.02	337.31
Shaw Crompton Way	0.86	-5.05	1292.77
Sheffield Barnsley Road	0.84	-5.62	2619.63
Sheffield Devonshire Green	0.79	-3.93	925.34
Sheffield Tinsley	0.84	-4.24	1046.24
Southampton A33	0.84	-9.85	3065.03
Southampton Centre	0.81	-4.78	1144.72
Southend-on-Sea	0.82	-2.17	349.35
Southwark A2 Old Kent Road	0.87	-8.39	3056.93
St Helens Linkway	0.85	-6.24	1631.40
St Osyth	0.80	-1.56	99.08
Stanford-le-Hope Roadside	0.88	-4.64	985.03
Stockton-on-Tees A1305 Roadside	0.81	-3.64	584.49
Stockton-on-Tees Eaglescliffe	0.78	-3.59	495.18
Stoke-on-Trent A50 Roadside	0.89	-7.54	3630.46
Stoke-on-Trent Centre	0.87	-3.10	518.29
Storrington Roadside	0.85	-3.92	938.07
Sunderland Silksworth	0.82	-2.04	231.45
Sunderland Wessington Way	0.84	-3.94	648.74
Thurrock	0.87	-4.64	1092.48
Tower Hamlets Roadside	0.88	-4.59	1666.94
Walsall Woodlands	0.85	-2.31	402.85
Warrington	0.81	-3.57	631.86
Wicken Fen	0.83	-1.06	48.99
Widnes Milton Road	0.85	-8.91	2756.66
Wigan Centre	0.86	-2.75	463.44
Wirral Tranmere	0.83	-2.12	251.85
Worthing A27 Roadside	0.88	-3.26	1056.68
Yarner Wood	0.72	-0.56	12.06
York Bootham	0.80	-2.19	316.08
York Fishergate	0.88	-3.28	777.95

(b) **Mean, Max and Minimum values for Bias, Correlation and MSE for NO_X across all air pollution monitoring stations. (cont.)**

Supplementary Table S10

Monitoring Station	Correlation	Bias (µg/m^3)	MSE (µg/m^3)^2
Aston Hill	0.59	-0.07	0.21
Barnsley Gawber	0.74	-1.41	122.92
Billingham	0.64	-2.12	158.75
Birkenhead Borough Road	0.74	-1.99	189.94
Birmingham A4540 Roadside	0.79	-4.01	642.43
Birmingham Acocks Green	0.79	-1.45	119.08
Blackburn Accrington Road	0.76	-3.19	304.31
Blackpool Marton	0.56	-0.92	48.35
Bournemouth	0.70	-1.20	70.91
Bradford Mayo Avenue	0.84	-2.84	756.75
Brighton Preston Park	0.59	-2.70	251.61
Bristol St Paul's	0.77	-4.18	599.12
Bury Whitefield Roadside	0.76	-3.23	350.71
Cambridge Roadside	0.84	-4.15	483.45
Camden Kerbside	0.85	-5.10	1372.93
Cannock A5190 Roadside	0.78	-1.49	119.02
Canterbury	0.67	-1.24	70.21
Carlisle Roadside	0.77	-3.76	511.78
Charlton Mackrell	0.56	-0.26	4.08
Chatham Roadside	0.79	-2.06	130.05
Chesterfield Loundsley Green	0.75	-1.98	119.85
Chesterfield Roadside	0.78	-2.03	172.28
Chilbolton Observatory	0.65	-0.85	32.40
Christchurch Barrack Road	0.70	-4.78	472.31
Coventry Allesley	0.79	-2.08	163.89
Doncaster A630 Cleveland Street	0.78	-2.10	225.24
Eastbourne	0.54	-1.18	72.53
Exeter Roadside	0.80	-5.07	641.91
Glazebury	0.68	-2.40	172.99
Haringey Roadside	0.83	-3.68	591.27
High Muffles	0.45	-0.14	1.15
Honiton	0.65	-0.39	14.39
Horley	0.77	-2.58	231.67
Hull Freetown	0.79	-1.37	99.92
Hull Holderness Road	0.81	-2.99	297.22
Ladybower	0.72	-0.25	2.21
Leamington Spa	0.67	-1.35	110.50
Leamington Spa Rugby Road	0.80	-1.58	124.88
Leeds Centre	0.73	-2.80	312.35
Leeds Headingley Kerbside	0.84	-2.65	469.26
Leicester A594 Roadside	0.84	-3.34	420.53
Leicester University	0.79	-2.20	163.87
Leominster	0.65	-0.76	34.45
Liverpool Speke	0.68	-1.76	126.53
London Bexley	0.79	-2.88	330.56
London Bloomsbury	0.78	-2.78	301.98
London Eltham	0.81	-1.67	137.05
London Haringey Priory Park South	0.79	-2.16	237.99
London Harlington	0.83	-3.67	506.72
London Hillingdon	0.83	-4.18	967.28
London Marylebone Road	0.88	-4.27	2925.59

(a) **Mean, Max and Minimum values for Bias, Correlation and MSE for NO across all air pollution monitoring stations.**

Monitoring Station	Correlation	Bias (µg/m^3)	MSE (µg/m^3)^2
London N. Kensington	0.80	-2.95	411.51
London Westminster	0.83	-2.07	192.27
Lullington Heath	0.49	-0.15	1.48
Luton A505 Roadside	0.84	-4.81	1263.03
Manchester Piccadilly	0.77	-3.78	397.84
Manchester Sharston	0.63	-3.30	458.55
Middlesbrough	0.63	-1.73	148.66
Newcastle Centre	0.72	-1.80	200.53
Newcastle Cradlewell Roadside	0.79	-6.12	996.23
Norwich Lakenfields	0.75	-0.79	43.44
Nottingham Centre	0.73	-2.24	302.73
Nottingham Western Boulevard	0.82	-4.75	679.25
Oldbury Birmingham Road	0.76	-3.66	785.25
Oxford Centre Roadside	0.84	-5.82	1150.00
Oxford St Ebbes	0.75	-2.83	256.82
Plymouth Centre	0.65	-3.20	290.81
Preston	0.72	-1.24	90.57
Reading New Town	0.78	-2.29	268.97
Rochester Stoke	0.63	-0.99	52.48
Salford Eccles	0.72	-3.31	453.22
Sandy Roadside	0.76	-2.58	424.30
Scunthorpe Town	0.72	-1.08	94.87
Shaw Crompton Way	0.80	-3.49	422.00
Sheffield Barnsley Road	0.76	-4.26	986.12
Sheffield Devonshire Green	0.66	-2.94	310.06
Sheffield Tinsley	0.76	-3.01	368.99
Southampton A33	0.75	-8.03	1180.07
Southampton Centre	0.74	-3.44	428.14
Southend-on-Sea	0.72	-1.42	116.45
Southwark A2 Old Kent Road	0.80	-6.38	1153.36
St Helens Linkway	0.79	-4.01	550.32
St Osyth	0.60	-0.65	19.13
Stanford-le-Hope Roadside	0.81	-3.44	395.10
Stockton-on-Tees A1305 Roadside	0.68	-2.33	173.81
Stockton-on-Tees Eaglescliffe	0.65	-1.97	130.88
Stoke-on-Trent A50 Roadside	0.83	-4.49	1341.83
Stoke-on-Trent Centre	0.82	-2.26	157.40
Storrington Roadside	0.77	-3.05	359.01
Sunderland Silksworth	0.57	-0.38	89.30
Sunderland Wessington Way	0.77	-2.81	190.53
Thurrock	0.83	-3.81	439.98
Tower Hamlets Roadside	0.82	-3.63	625.32
Walsall Woodlands	0.77	-1.71	158.00
Warrington	0.72	-2.56	242.80
Wicken Fen	0.62	-0.48	12.98
Widnes Milton Road	0.78	-6.72	1014.43
Wigan Centre	0.77	-2.24	179.52
Wirral Tranmere	0.70	-1.16	66.04
Worthing A27 Roadside	0.84	-3.19	325.35
Yarner Wood	0.47	-0.11	0.55
York Bootham	0.56	-2.07	139.27
York Fishergate	0.83	-2.41	280.54

(b) **Mean, Max and Minimum values for Bias, Correlation and MSE for NO across all air pollution monitoring stations. (cont.)**

Supplementary Table S11

Monitoring Station	Correlation	Bias (μg/m^3)	MSE (μg/m^3)^2
Barnstaple A39	0.70	-1.60	60.28
Birmingham A4540 Roadside	0.76	-1.47	57.07
Bristol St Paul's	0.74	-1.54	61.49
Bury Whitefield Roadside	0.75	-1.35	51.80
Camden Kerbside	0.82	-1.39	67.12
Carlisle Roadside	0.36	-1.92	335.28
Chatham Roadside	0.77	-1.56	69.23
Chesterfield Loundsley Green	0.78	-1.24	39.50
Chesterfield Roadside	0.77	-1.43	66.32
Chilbolton Observatory	0.70	-1.55	48.70
Ealing Horn Lane	0.75	-1.97	156.77
Hull Holderness Road	0.70	-1.88	94.31
Leamington Spa	0.43	-1.34	253.61
Leamington Spa Rugby Road	0.80	-1.16	39.37
Leeds Centre	0.78	-1.58	63.01
Leeds Headingley Kerbside	0.76	-1.67	92.75
Leicester A594 Roadside	0.79	-1.21	57.22
Liverpool Speke	0.70	-1.58	71.12
London Bloomsbury	0.80	-1.40	58.65
London Harlington	0.79	-1.42	56.64
London Marylebone Road	0.81	-1.27	64.00
London N. Kensington	0.82	-1.29	54.13
Middlesbrough	0.69	-1.71	80.93

Supplementary Table S12: **Mean, Max and Minimum values for Bias, Correlation and MSE for PM_{10} across all air pollution monitoring stations.**

Monitoring Station	Correlation	Bias ($\mu g/m^3$)	MSE $(\mu g/m^3)^2$
Barnstaple A39	0.72	-1.48	30.07
Birmingham A4540 Roadside	0.80	-1.30	35.95
Birmingham Acocks Green	0.80	-1.17	34.23
Blackpool Marton	0.67	-1.87	52.07
Bristol St Paul's	0.76	-1.54	52.64
Camden Kerbside	0.82	-1.23	42.55
Carlisle Roadside	0.71	-1.34	36.63
Chatham Roadside	0.76	-1.50	50.57
Chesterfield Loundsley Green	0.75	-1.23	31.18
Chesterfield Roadside	0.74	-1.39	50.38
Chilbolton Observatory	0.71	-1.57	36.26
Coventry Allesley	0.78	-1.24	36.44
Eastbourne	0.71	-1.41	47.83
Hull Freetown	0.74	-1.45	46.61
Leamington Spa	0.79	-1.15	33.40
Leamington Spa Rugby Road	0.80	-1.23	31.95
Leeds Centre	0.76	-1.36	53.43
Leeds Headingley Kerbside	0.76	-1.21	52.17
Leicester University	0.78	-1.15	37.96
Liverpool Speke	0.71	-1.34	46.68
London Bexley	0.81	-1.13	38.14
London Bloomsbury	0.80	-1.47	43.95
London Eltham	0.79	-1.41	39.58
London Harlington	0.81	-1.17	37.39
London Marylebone Road	0.79	-1.22	45.05
London N. Kensington	0.81	-1.38	44.91
London Teddington Bushy Park	0.79	-1.58	41.82
Manchester Piccadilly	0.67	-1.82	68.72
Middlesbrough	0.71	-1.69	50.50
Newcastle Centre	0.77	-1.33	30.23
Norwich Lakenfields	0.73	-1.39	45.09
Nottingham Centre	0.76	-1.49	48.41
Oxford St Ebbes	0.79	-1.10	26.51
Plymouth Centre	0.66	-1.39	50.39
Preston	0.72	-1.13	42.45
Reading New Town	0.79	-1.86	41.20
Rochester Stoke	0.74	-1.91	53.06
Salford Eccles	0.68	-1.18	73.46
Saltash Callington Road	0.70	-1.51	31.04
Sandy Roadside	0.79	-1.29	40.56
Sheffield Devonshire Green	0.75	-1.64	51.49
Southampton Centre	0.75	-1.32	45.01
Southend-on-Sea	0.79	-1.52	49.38
Stanford-le-Hope Roadside	0.78	-1.93	49.61
Stockton-on-Tees A1305 Roadside	0.73	-1.76	45.88
Stockton-on-Tees Eaglescliffe	0.69	-2.39	59.44
Stoke-on-Trent Centre	0.72	-1.53	48.12
Sunderland Silksworth	0.72	-1.39	34.38
Warrington	0.78	-1.14	37.44
Wigan Centre	0.72	-1.49	65.69
Wirral Tranmere	0.76	-1.74	36.13
York Bootham	0.76	-1.55	41.79
York Fishergate	0.78	-1.33	41.81

Supplementary Table S13: **Mean, Max and Minimum values for Bias, Correlation and MSE for $PM_{2.5}$ across all air pollution monitoring stations.**

Monitoring Station	Correlation	Bias (µg/m^3)	MSE (µg/m^3)^2
Barnsley Gawber	0.72	-0.27	1.80
Chilbolton Observatory	0.53	-0.14	0.25
Hull Freetown	0.52	-0.36	2.33
Ladybower	0.53	-0.32	1.77
Leeds Centre	0.62	-0.21	1.71
Liverpool Speke	0.45	-0.70	12.23
London Bloomsbury	0.48	-0.41	3.63
London Marylebone Road	0.88	-0.89	5.54
London N. Kensington	0.56	-0.33	2.56
Lullington Heath	0.35	-0.25	1.99
Manchester Piccadilly	0.61	-0.25	1.39
Middlesbrough	0.58	-0.83	10.52
Nottingham Centre	0.51	-0.36	2.20
Rochester Stoke	0.57	-0.43	1.61
Scunthorpe Town	0.68	-1.45	30.77
Southampton Centre	0.57	-0.34	2.32
Thurrock	0.62	-0.26	1.89
Wicken Fen	0.61	-0.33	2.84

Supplementary Table S14: **Mean, Max and Minimum values for Bias, Correlation and MSE for SO_2 across all air pollution monitoring stations.**

Pollutant Name	Dataset Train Score	Dataset Validation Score	Dataset Test Score
NO_2	-0.0	0.03	0.02
NO_x	-0.02	-0.01	0.01
NO	-0.06	-0.06	-0.05
O_3	0.12	0.04	0.0
SO_2	-0.06	0.0	0.0
PM_{10}	-0.01	-0.01	-0.05
PM_{25}	-0.05	-0.03	-0.08

(a) **Temporal**

Pollutant Name	Dataset Train Score	Dataset Validation Score	Dataset Test Score
NO_2	0.14	0.16	0.14
NO_x	0.09	0.1	0.09
NO	0.03	0.02	0.01
O_3	0.43	0.41	0.34
SO_2	-0.02	0.04	0.03
PM_{10}	0.2	0.23	0.18
PM_{25}	0.25	0.26	0.24

(b) **Meteorology**

Pollutant Name	Dataset Train Score	Dataset Validation Score	Dataset Test Score
NO_2	0.32	0.29	0.31
NO_x	0.26	0.25	0.27
NO	0.19	0.18	0.2
O_3	-0.02	-0.04	-0.15
SO_2	0.16	0.26	0.12
PM_{10}	-0.01	0.02	-0.04
PM_{25}	-0.07	-0.02	-0.1

(c) **Transport Infrastructure**

Pollutant Name	Dataset Train Score	Dataset Validation Score	Dataset Test Score
NO_2	0.36	0.23	0.16
NO_x	0.28	0.18	0.14
NO	0.21	0.12	0.08
O_3	0.09	0.02	-0.09
SO_2	0.19	0.24	0.12
PM_{10}	0.02	0.03	-0.06
PM_{25}	-0.07	-0.03	-0.08

(d) **Transport Use**

Pollutant Name	Dataset Train Score	Dataset Validation Score	Dataset Test Score
NO_2	0.35	0.31	0.36
NO_x	0.28	0.25	0.32
NO	0.2	0.18	0.24
O_3	0.14	0.11	0.05
SO_2	0.18	0.26	0.13
PM_{10}	0.01	0.01	-0.01
PM_{25}	-0.05	-0.03	-0.05

(e) **Remote Sensing**

Pollutant Name	Dataset Train Score	Dataset Validation Score	Dataset Test Score
NO_2	0.31	0.23	0.29
NO_x	0.25	0.2	0.26
NO	0.18	0.12	0.2
O_3	-0.04	-0.1	-0.2
SO_2	0.17	0.23	0.13
PM_{10}	-0.01	0.02	-0.07
PM_{25}	-0.08	-0.03	-0.11

(f) **Geographic**

Pollutant Name	Dataset Train Score	Dataset Validation Score	Dataset Test Score
NO_2	0.47	0.42	0.43
NO_x	0.39	0.34	0.39
NO	0.32	0.28	0.28
O_3	0.1	-0.02	-0.01
SO_2	0.23	0.29	0.13
PM_{10}	0.05	0.06	-0.02
PM_{25}	-0.02	-0.0	-0.08

(g) **Emissions**

Supplementary Table S15: **Temporal experiment results for different subsets of dataset families.** All dataset families provide some information about at least one of the air pollutants within the study. Still, none provide the complete picture of air pollution concentrations in the study, aligning with the scientific literature concerning air pollution sources and sinks with various phenomena from meteorological conditions, such as wind, and emissions sources, such as transportation and industry impacting seen air pollution concentrations. As such, it is clear that all of the datasets outlined are needed to predict air pollution concentrations in the future at an adequate level and no single dataset.

S3.3 Data Subsetting - Spatial

Pollutant Name	Estimation LOOV Max	Estimation LOOV Min	Estimation LOOV Mean	Estimation LOOV Median
NO	0.11	-110.87	-2.16	-0.06
NO_2	0.29	-13.84	-0.63	-0.06
NO_x	0.22	-30.30	-1.09	-0.04
O_3	0.37	-2.44	-0.11	0.11
PM_{10}	0.05	-0.60	-0.06	0.01
PM_{25}	0.08	-0.60	-0.06	-0.06
SO_2	0.01	-1.97	-0.28	-0.07

(a) **Temporal**

Pollutant Name	Estimation LOOV Max	Estimation LOOV Min	Estimation LOOV Mean	Estimation LOOV Median
NO	0.35	-441.56	-5.97	0.08
NO_2	0.46	-14.89	-0.52	0.14
NO_x	0.44	-61.68	-1.49	0.13
O_3	0.70	-2.96	0.25	0.47
PM_{10}	0.40	-0.34	0.20	0.24
PM_{25}	0.38	-0.20	0.25	0.27
SO_2	0.08	-3.15	-0.30	-0.03

(b) **Meteorology**

Pollutant Name	Estimation LOOV Max	Estimation LOOV Min	Estimation LOOV Mean	Estimation LOOV Median
NO	0.00	-1.33	-0.20	-0.07
NO_2	0.00	-3.02	-0.41	-0.23
NO_x	0.00	-3.35	-0.31	-0.11
O_3	0.00	-1.39	-0.33	-0.20
PM_{10}	-0.00	-0.79	-0.15	-0.12
PM_{25}	0.00	-0.52	-0.11	-0.09
SO_2	0.00	-1.86	-0.22	-0.07

(c) **Transport Infrastructure**

Pollutant Name	Estimation LOOV Max	Estimation LOOV Min	Estimation LOOV Mean	Estimation LOOV Median
NO	0.03	-18.60	-0.59	-0.14
NO_2	0.01	-5.36	-0.46	-0.27
NO_x	0.03	-11.08	-0.50	-0.18
O_3	0.04	-3.08	-0.30	-0.22
PM_{10}	-0.01	-0.58	-0.14	-0.09
PM_{25}	-0.01	-0.42	-0.12	-0.12
SO_2	-0.01	-1.73	-0.26	-0.11

(d) **Transport Use**

Pollutant Name	Estimation LOOV Max	Estimation LOOV Min	Estimation LOOV Mean	Estimation LOOV Median
NO	0.01	-35.53	-0.80	-0.10
NO_2	0.05	-2.41	-0.38	-0.21
NO_x	0.03	-6.96	-0.44	-0.21
O_3	0.20	-1.54	-0.16	-0.10
PM_{10}	0.03	-0.62	-0.10	-0.07
PM_{25}	0.02	-0.50	-0.10	-0.09
SO_2	-0.04	-1.79	-0.41	-0.21

(e) **Remote Sensing**

Pollutant Name	Estimation LOOV Max	Estimation LOOV Min	Estimation LOOV Mean	Estimation LOOV Median
NO	0.00	-3.53	-0.31	-0.15
NO_2	0.04	-3.82	-0.46	-0.25
NO_x	0.00	-4.98	-0.41	-0.22
O_3	-0.00	-1.28	-0.36	-0.23
PM_{10}	-0.00	-0.72	-0.13	-0.08
PM_{25}	-0.00	-0.53	-0.14	-0.11
SO_2	-0.00	-2.36	-0.28	-0.08

(f) **Geographic**

Pollutant Name	Estimation LOOV Max	Estimation LOOV Min	Estimation LOOV Mean	Estimation LOOV Median
NO	0.13	-3.30	-0.23	-0.10
NO_2	0.25	-2.45	-0.23	-0.10
NO_x	0.17	-3.29	-0.25	-0.11
O_3	0.15	-2.34	-0.26	-0.19
PM_{10}	0.04	-0.73	-0.10	-0.05
PM_{25}	0.02	-0.57	-0.09	-0.06
SO_2	0.02	-2.13	-0.31	-0.18

(g) **Emissions**

Supplementary Table S16: **Summary statistics for individual monitoring station leave-one-out-validation (LOOV) for the spatial experiment with different subsets of dataset families.** The spatial LOOV experiments echo the results seen in Table S15 to a more extreme degree, where the median for the majority of the LOOV is negative when only a single dataset family is included, highlighting that the prediction of the concentrations performs worse than simply predicting the average for the station's measurements, highlighting the importance of including a range of phenomena data to be able to accurately predict the air pollution concentrations of a monitoring stations concentration measurements.

S3.4 Data Subsetting - Forecasting and Global Framework

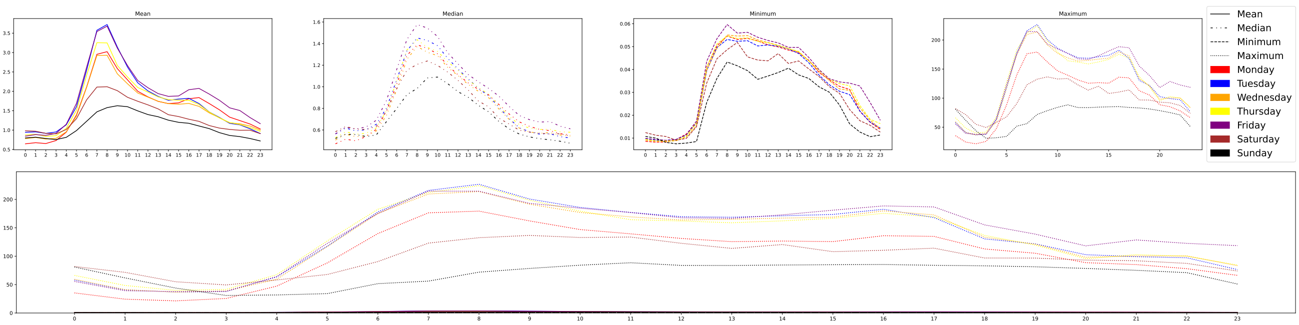
Pollutant Name	Dataset Train Score	Dataset Validation Score	Dataset Test Score	Mean LOOV
NO_2	0.76	0.68	0.70	-0.09
NO_x	0.73	0.65	0.68	-0.69
NO	0.46	0.43	0.40	-1.87
O_3	0.77	0.66	0.62	0.48
SO_2	0.28	0.37	0.25	-0.47
PM_{10}	0.49	0.35	0.31	0.38
PM_{25}	0.53	0.35	0.30	0.46

Supplementary Table S17: **Overview of forecasting (test score) and filling missing data spatial (LOOV) performance for a model trained with data that is only available globally.** When using the dataset families available globally, it can be seen that the model's performance in a forecasting situation is good. However, the LOOV is weaker, with some air pollutants having a negative performance overall, highlighting that the data that is available globally would be suitable for forecasting into the future in locations where monitoring stations are, but not when used to fill in missing monitoring station locations, without further improvements to the model or data used. The global datasets are the temporal, meteorological and remote sensing dataset families in this context.

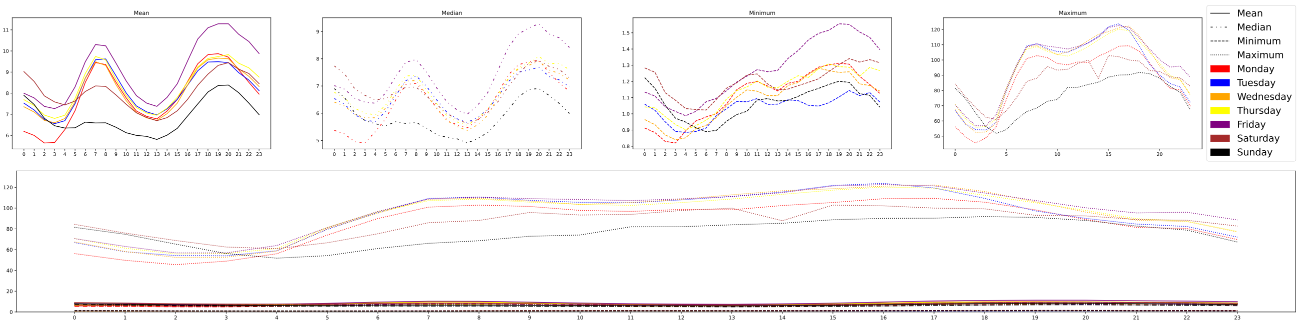
Pollutant Name	Dataset Train Score	Dataset Validation Score	Dataset Test Score	Mean LOOV
NO_2	0.83	0.77	0.76	0.11
NO_x	0.80	0.74	0.73	0.03
NO	0.69	0.63	0.59	-0.05
O_3	0.79	0.69	0.66	0.39
SO_2	0.37	0.38	0.25	-0.11
PM_{10}	0.51	0.38	0.33	0.32
PM_{25}	0.52	0.35	0.31	0.42

Supplementary Table S18: **Overview of forecasting (test score) and filling missing data spatial (LOOV) performance for a model trained with data that is only available ahead of time (e.g. forecasting ahead of the current date).** When using the dataset families available ahead of time that could be used in a true forecasting mode, e.g. predict ahead of the current date, it can be seen that the model's performance in a forecasting situation is similar to the global model seen in Table S17, showing good performance. However, the LOOV is considerably better, with most of the air pollutants having a strong LOOV performance, with only NO and SO_2 having negative performance, driven by a small number of bad predictions for particular stations. In this context, the forecasting datasets are road infrastructure, geographic, meteorological and temporal dataset families. The difference between the performance of the forecasting datasets and the global datasets models shows the importance of some datasets to the model's overall performance, such as the road infrastructure is critical for NO_x and NO_2 where its inclusion improves performance considerably, particularly for the LOOV results.

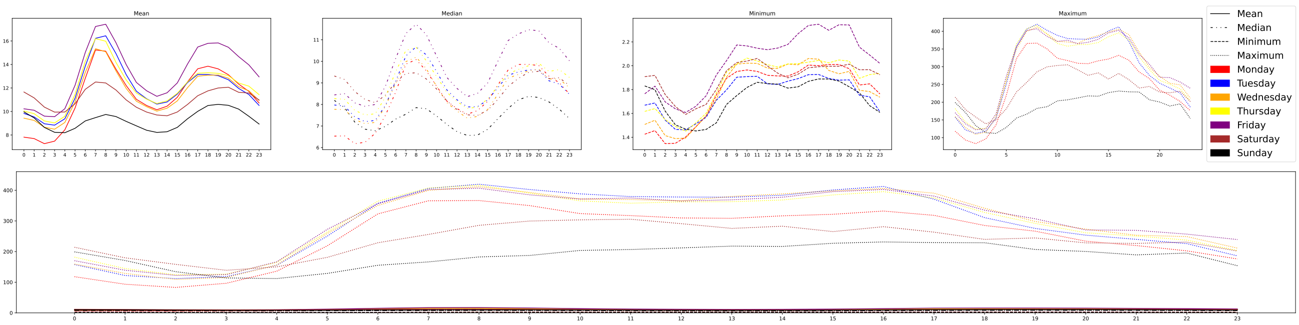
S4 Research Data Output Summary Statistics



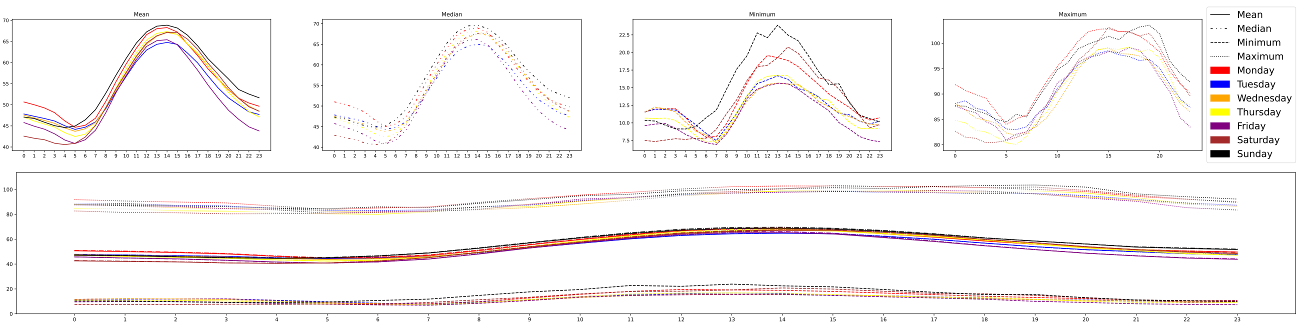
(a) NO



(b) NO2

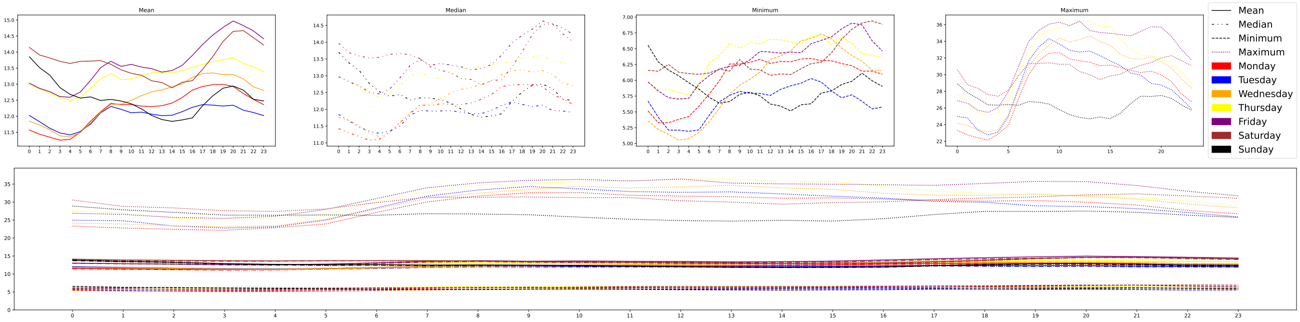


(c) NOX

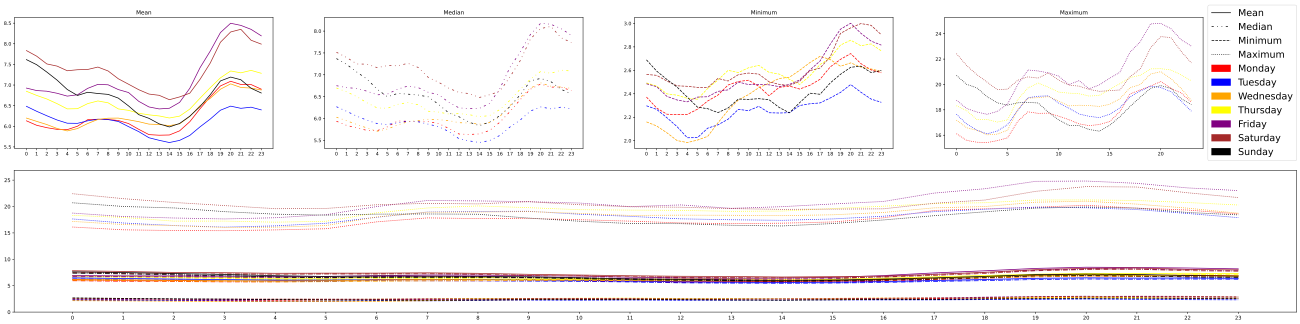


(d) O3

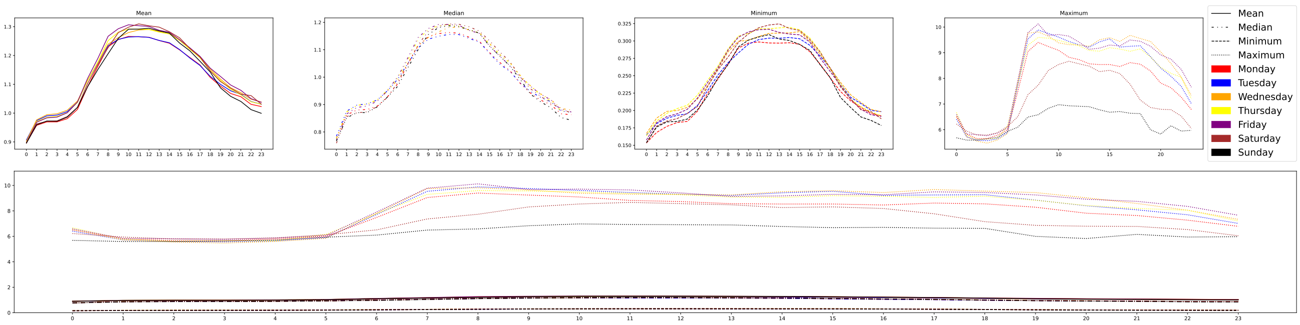
Supplementary Figure S21: **Summary of the complete air pollution concentration dataset, spatially and temporal for England at the 1km^2 spatial resolution for the air pollutants NO, NO_2, NO_x, O_3, PM_{10}, $PM_{2.5}$ and SO_2 at the hourly temporal level for 2018.**



(e) PM10



(f) $PM_{2.5}$



(g) SO2

Supplementary Figure S21: **Summary of the complete air pollution concentration dataset, spatially and temporal for England at the 1km^2 spatial resolution for the air pollutants NO, NO_2, NO_x, O_3, PM_{10}, $PM_{2.5}$ and SO_2 at the hourly temporal level for 2018. (cont.)** Summarised in the figures is the dataset made possible by the model developed and presented by this work. The mean, median, minimum and maximum at each hour of the day for each day of the week are shown for each air pollutant, highlighting their overall trends across England.

References

[1] Daewon W. Byun, Avraham Lacser, Robert Yamartino, and Paolo Zannetti, “Chapter 10 eulerian dispersion models,” 2003.

[2] D. W. Wong, “The modifiable areal unit problem (maup),” in *WorldMinds: geographical perspectives on 100 problems: commemorating the 100th anniversary of the association of American geographers 1904–2004*. Springer, 2004, pp. 571–575.

[3] [Online]. Available: https://uk-air.defra.gov.uk/data/pcm-data

[4] Department of Transport, UK Government. (2023) Road Traffic Stations - About. Accessed on: 29/11/2023. [Online]. Available: https://roadtraffic.dft.gov.uk/about

[5] ——. (2023) Road Traffic Statistics - Meta Data. Accessed on: 29/11/2023. [Online]. Available: https://storage.googleapis.com/dft-statistics/road-traffic/all-traffic-data-metadata.pdf

[6] J. Gershuny, “United kingdom time use survey, 2014-2015,” 2017.

[7] O. for National Statistics, “2011 census aggregate data,” 2017. [Online]. Available: https://www.ons.gov.uk/census/2011census

[8] R. Lovelace and M. Dumont, *Spatial microsimulation with R*. Chapman and Hall/CRC, 2017.

[9] R. L. Dumont and Morgane, “Spatial microsimulation with r,” May 2018. [Online]. Available: https://spatial-microsim-book.robinlovelace.net/svalidation#Internal

[10] H. Hersbach, “The era5 atmospheric reanalysis.” in *AGU fall meeting abstracts*, vol. 2016, 2016, pp. NG33D–01.

[11] N. Gorelick, “Google earth engine,” in *EGU general assembly conference abstracts*, vol. 15. American Geophysical Union Vienna, Austria, 2013, p. 11997.

[12] J. P. Veefkind, I. Aben, K. McMullan, H. Förster, J. De Vries, G. Otter, J. Claas, H. Eskes, J. De Haan, Q. Kleipool *et al.*, “Tropomi on the esa sentinel-5 precursor: A gmes mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications,” *Remote sensing of environment*, vol. 120, pp. 70–83, 2012.

[13] UK National Atmospheric Emissions Inventory (NAEI). (2023) NAEI - Data. Accessed on: 29/11/2023. [Online]. Available: https://naei.beis.gov.uk/data/

[14] ——. (2023) UN/ECE Classification of Emission Sources. Accessed on: 29/11/2023. [Online]. Available: https://uk-air.defra.gov.uk/assets/documents/reports/empire/naei/annreport/annrep98/naeiapp4.html

[15] ——. (2023) Emissions from NAEI large point sources. Accessed on: 29/11/2023. [Online]. Available: https://naei.beis.gov.uk/data/map-large-source

[16] H. Denier van der Gon, C. Hendriks, J. Kuenen, A. Segers, and A. Visschedijk, “Description of current temporal emission patterns and sensitivity of predicted aq for temporal emission patterns,” *EU FP7 MACC deliverable report D_D-EMIS_1*, vol. 3, pp. 2019–07, 2011.

[17] C. Rowland, R. Morton, L. Carrasco, G. McShane, A. O’neil, and C. Wood, “Land cover map 2015 (vector, gb),” *NERC Environmental Information Data Centre*, vol. 10, 2017.