

SPF Clean Air

UK Air Quality Reanalysis

Dataset Documentation

Eleanor Smith, Lucy Neal, Paul Agnew

Version 1.1

Version History

Version Number	Date	Comments
1.0	30 th March 2023	Dataset documentation to accompany the initial release of the AQREAN reanalysis dataset
1.1	31 st January 2024	Updated documentation to reflect modifications and additions to the AQREAN reanalysis dataset at version 1.1

Contents

Version History.....	2
Contents	3
1. Overview.....	4
2. The AQREAN Model.....	5
2.1. Model Configuration.....	5
2.2. Boundary Conditions.....	6
2.3. Emission Inventories.....	6
2.4. Production Simulations	6
3. Post Processing	7
3.1. Bias Correction Methodology	7
3.2. Observations.....	8
3.2.1. Ozone (O ₃).....	9
3.2.2. Nitric Oxide (NO).....	10
3.2.3. Nitrogen Dioxide (NO ₂)	11
3.2.4. Sulphur Dioxide (SO ₂).....	12
3.2.5. Carbon Monoxide (CO).....	13
3.2.6. Fine Particulate Matter (PM _{2.5})	14
3.2.7. Coarse Particulate Matter (PM ₁₀)	15
4. Dataset Contents	16
4.1. Dataset Domain	16
4.2. Surface Level Data	17
4.2.1. Data Description	17
4.2.2. Main Variables	17
4.3. Bias Corrected Surface Level Data	20
4.3.1. Data Description	20
4.3.2. Main Variables	20
5. Acronyms.....	22
6. References	24

1. Overview

The UK Air Quality Reanalysis has been generated as part of the SPF Clean Air Programme.

The dataset contains the concentrations of gaseous and particulate air pollutants as well as the daily air quality index and a selection of meteorological variables. The data covers the UK at a horizontal resolution of 0.1 degree (approximately 10 km) and is available at hourly time frequency for a 17 year period from 1st January 2003 to 31st December 2019 inclusively.

Further details of this dataset are given in the sections below.

2. The AQREAN Model

The Met Office generate the UK national air quality forecast using Air Quality in the Unified Model (AQUM) which is an online, limited-area forecast configuration of the Unified Model (UM) ([Savage et al., 2013](#)). The reanalysis model AQREAN has been developed based on this existing operational air quality forecast model.

2.1. Model Configuration

The AQREAN model suite has been developed by incorporating the desired features from both the European domain AQUM model suite (AQCOPS) and the UK domain AQUM model suite (AQEUR). For more details of the AQUM model set up see [Savage et al., 2013](#).

The horizontal domain of the AQREAN model covers the UK as well as parts of North West Europe at a resolution of 0.11 degrees (approximately 12 km). AQREAN uses a rotated pole grid with the pole coordinates at 177.5 °E and 37.5 °N. The model has a vertical resolution of 63 levels which extend from the surface to approximately 40 km, with the lowest model level at 20 m. The horizontal domain of AQREAN is illustrated in Figure 2.1.

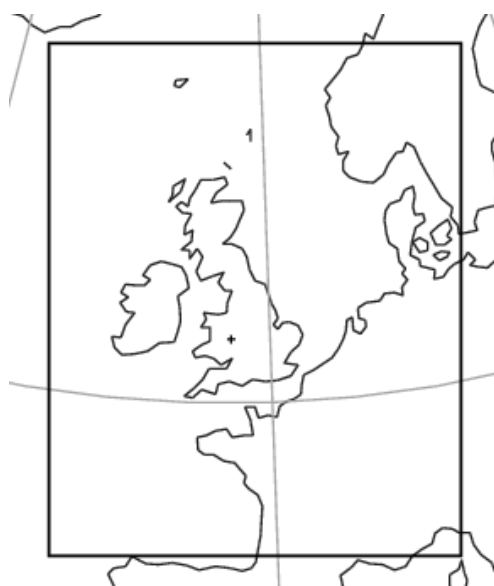


Figure 2.1: Map illustrating the horizontal domain of the AQREAN model.

AQREAN uses the UKCA (United Kingdom Chemistry and Aerosols) chemical scheme with the Regional Air Quality (RAQ) chemical mechanism which includes 40 chemical species and over 100 chemical reactions. The Coupled Large-scale Aerosol Simulator for Studies in Climate (CLASSIC) aerosol module is used in AQREAN and includes black carbon, organic carbon, sulphate, nitrate, dust and secondary organic aerosols.

AQREAN runs with a model timestep of 5 minutes. The simulations begin at 00Z and are run as 24 hour forecasts.

2.2. Boundary Conditions

Meteorological and chemical composition data is required to generate initial fields as well as lateral boundary condition (LBC) files for running AQREAN. Datasets from the European Centre for Medium-Range Weather Forecasts (ECMWF) are used to provide this data.

The ECMWF 5th Generation Reanalysis (ERA5) dataset has been chosen to provide the meteorological data and the ECMWF 4th Generation Atmospheric Composition Reanalysis (EAC4) dataset has been chosen to provide the chemical composition data.

Ozone LBCs have been shown to have significant biases in the operational AQUM model which are corrected prior to the model being run. This method is also applied to the AQREAN simulations and the ozone concentrations in the boundary conditions are adjusted using monthly bias corrections.

2.3. Emission Inventories

Pollutant emissions occur from a range of different sources. The AQREAN simulations require emissions data from anthropogenic, biogenic and biomass burning sources.

The Copernicus Atmosphere Monitoring Service (CAMS) regional European emissions (CAMS-REG-AP, <https://permalink.aeris-data.fr/CAMS-REG-AP>) has been selected for anthropogenic emissions.

The Global Fire Assimilation System (GFAS, <https://www.ecmwf.int/en/forecasts/dataset/global-fire-assimilation-system>) inventory has been selected to provide biomass burning emissions. Estimations of the altitude of the plume bottom are made before July 2018 where this information is not available in the dataset.

The CAMS-GLOB-BIO (<https://permalink.aeris-data.fr/CAMS-GLOB-BIO>) global biogenic emission inventory produced by ECMWF has been selected to provide biogenic isoprene emissions.

2.4. Production Simulations

AQREAN simulations have been run for 17 years from 2003 to 2019 inclusively. Each year is run separately, with a 1 month spin up period from 1st December the previous year. Note that the 1 month spin up period uses emissions data from the production year rather than the emissions data from the previous year.

3. Post Processing

A Statistical Post Processing of Observations (SPPO) method is applied to the AQUM operational air quality forecasts and is described in full in [Neal et al., 2014](#). This SPPO process has been adapted for use with the AQREAN model data and is used to produce the reanalysed surface level concentration data.

3.1. Bias Correction Methodology

For the operational air quality forecasts, observations of pollutants from the recent past are used to estimate current model biases. However, for the reanalysis, the model is run over past time periods and so it is possible to derive actual biases using model and observational data at the same time.

The SPPO process is applied to the model concentrations of ozone (O₃), nitric oxide (NO), nitrogen dioxide (NO₂), sulphur dioxide (SO₂), carbon monoxide (CO), particulate matter with diameter less than 2.5 µm (PM_{2.5}) and particulate matter with diameter less than 10 µm (PM₁₀).

As only surface level observations are used in the post processing, the bias corrections are only applied to the surface level model data and no post processing of pollutant concentrations for higher model levels is performed.

At ~12 km horizontal resolution, the AQREAN model is relatively coarse and so cannot accurately represent pollutant concentrations in locations that are close to strong emission sources such as roadsides and industrial sites. The bias corrections therefore only use observational data from locations which are classified as “background” locations (“rural background”, “suburban background” and “urban background”). These background sites are defined as being located such that the pollution level is not significantly influenced by any single source and the sampling point should be representative of several square kilometres (<https://uk-air.defra.gov.uk/networks/site-types>).

The majority of background measurement sites are in urban locations and this can cause some issues with the bias corrections in remote areas where there are few observations. This is particularly true for nitrogen oxides (NO and NO₂) over Northern Scotland, Wales and South West England. These are typically rural areas. However, the bias corrections are determined largely from urban background observations which means that the post processing method results in these concentrations being adjusted to higher concentrations (more typical of urban areas) across these remote regions.

The SPPO method used for the reanalysis dataset uses a weighting method based upon emission maps to adjust the contribution from the rural and urban sites and provide a better adjustment to the concentrations, particularly over the remote rural regions with sparse observations.

3.2. Observations

In order to perform the bias corrections to generate the reanalysed data, ground based observations of pollutants are required. These observations are largely obtained from the Automatic Urban and Rural Network (AURN, <https://uk-air.defra.gov.uk/networks/network-info?view=aurn>) data archive, with additional observation sites from the Republic of Ireland incorporated as well. Data for the sites in the Republic of Ireland has been retrieved through openair ([Carslaw et al., 2012](#)). Observations of O₃, NO, NO₂, SO₂, CO, PM_{2.5} and PM₁₀ from these observation sites are used in the post processing.

The availability of background observations for each of the species included is summarised in the sections below.

3.2.1. Ozone (O_3)

Table 3.1: Number of background sites (urban, suburban, rural and total) which measure O_3 and are included in the reanalysis post processing for each of the years.

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Urban	55	57	55	55	54	46	46	45	46	48	50	48	45	47	45	45	46
Suburban	4	4	6	8	8	6	8	10	10	10	10	8	9	10	10	10	10
Rural	27	26	28	29	29	30	29	28	28	27	27	26	27	28	25	27	27
Total	86	87	89	92	91	82	83	83	84	85	87	82	81	85	80	82	83

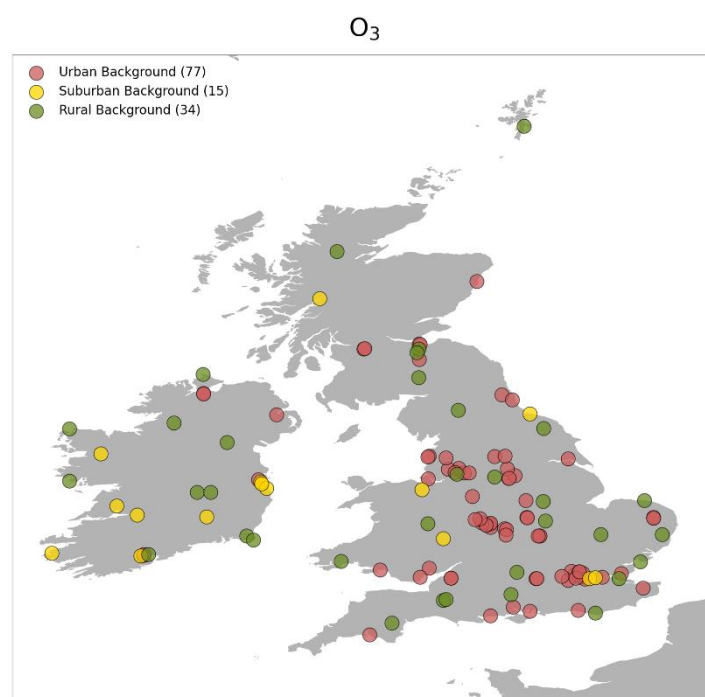


Figure 3.1: Map illustrating the locations of the background sites which measure O_3 during the current reanalysis period (2003 – 2019).

3.2.2. Nitric Oxide (NO)

Table 3.2: Number of background sites (urban, suburban, rural and total) which measure NO and are included in the reanalysis post processing for each of the years.

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Urban	61	63	61	61	60	52	52	51	51	53	54	53	51	55	58	60	62
Suburban	3	3	4	5	5	4	4	5	5	5	5	5	4	4	4	4	4
Rural	13	15	15	15	15	16	15	15	15	15	15	15	15	16	15	15	15
Total	77	81	80	81	80	72	71	71	71	73	74	73	70	75	77	79	81

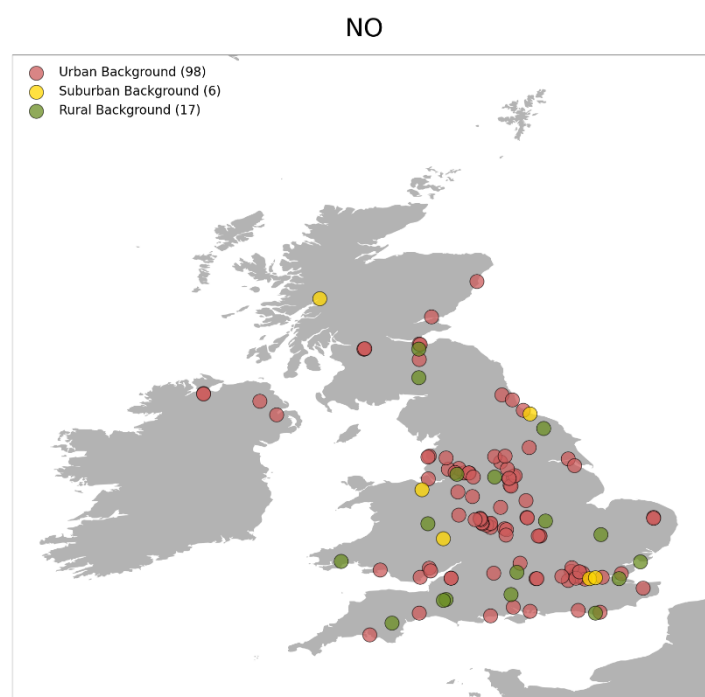


Figure 3.2: Map illustrating the locations of the background sites which measure NO during the current reanalysis period (2003 – 2019).

3.2.3. Nitrogen Dioxide (NO₂)

Table 3.3: Number of background sites (urban, suburban, rural and total) which measure NO₂ and are included in the reanalysis post processing for each of the years.

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Urban	63	65	63	63	62	53	55	54	54	56	57	55	52	58	61	64	66
Suburban	4	6	6	9	9	9	10	12	11	12	11	11	12	11	10	11	11
Rural	15	17	17	18	18	17	17	17	17	17	17	17	17	18	17	17	17
Total	82	88	86	90	89	79	82	83	82	85	85	83	81	87	88	92	94

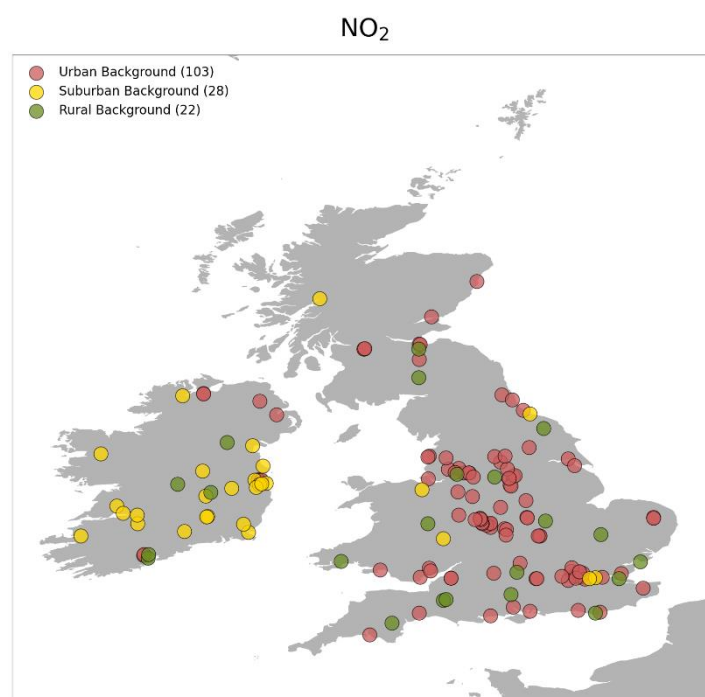


Figure 3.3: Map illustrating the locations of the background sites which measure NO₂ during the current reanalysis period (2003 – 2019).

3.2.4. Sulphur Dioxide (SO₂)

Table 3.4: Number of background sites (urban, suburban, rural and total) which measure SO₂ and are included in the reanalysis post processing for each of the years.

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Urban	57	56	54	55	54	29	30	31	32	32	21	17	16	19	18	17	16
Suburban	3	6	5	6	5	5	7	7	6	7	4	6	6	5	4	5	5
Rural	7	7	7	8	8	7	7	7	7	7	7	7	7	8	7	7	7
Total	67	69	66	69	67	41	44	45	45	46	32	30	29	32	29	29	28

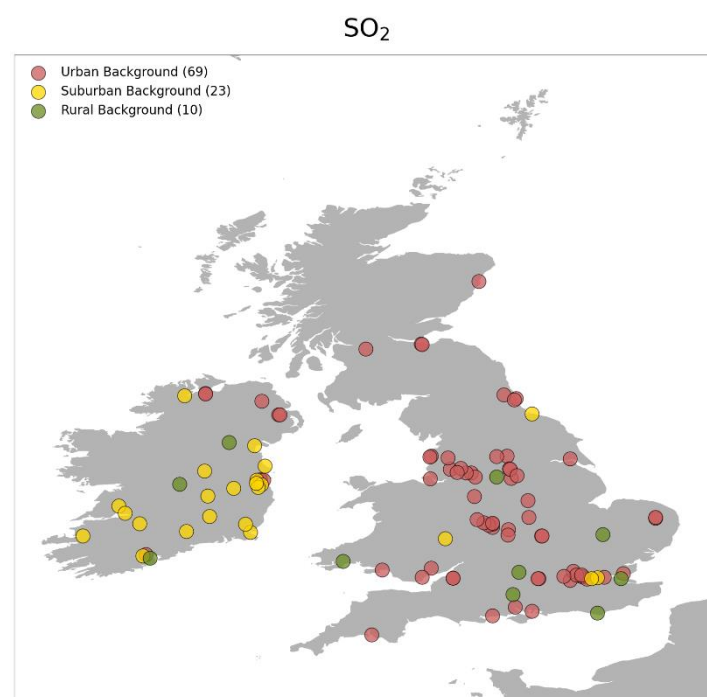


Figure 3.4: Map illustrating the locations of the background sites which measure SO₂ during the current reanalysis period (2003 – 2019).

3.2.5. Carbon Monoxide (CO)

Table 3.5: Number of background sites (urban, suburban, rural and total) which measure CO and are included in the reanalysis post processing for each of the years.

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Urban	55	55	53	53	52	16	17	17	17	17	8	6	5	7	7	7	7
Suburban	2	4	3	3	2	3	3	3	3	4	2	3	3	2	1	2	2
Rural	2	2	2	3	4	2	2	2	0	0	0	0	0	0	0	0	0
Total	59	61	58	59	58	21	22	22	20	21	10	9	8	9	8	9	9

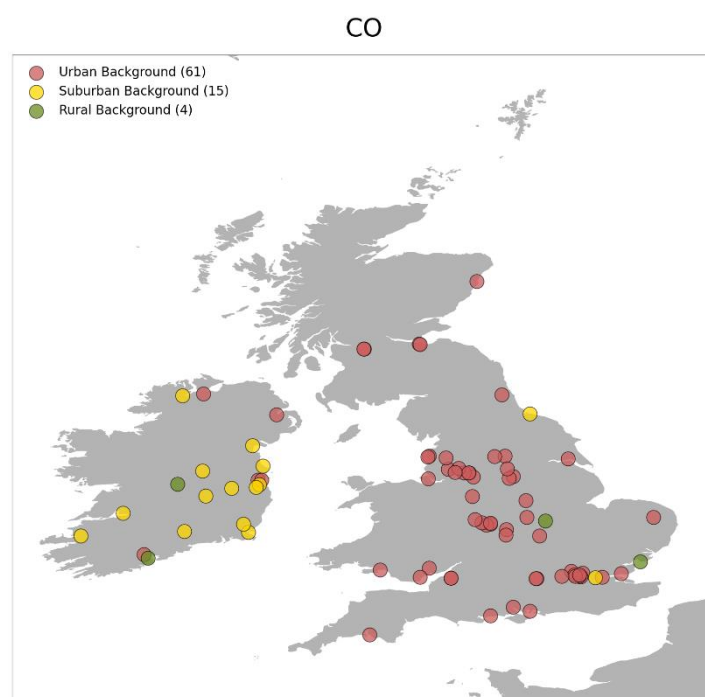


Figure 3.5: Map illustrating the locations of the background sites which measure CO during the current reanalysis period (2003 – 2019).

3.2.6. Fine Particulate Matter ($PM_{2.5}$)

Table 3.6: Number of background sites (urban, suburban, rural and total) which measure $PM_{2.5}$ and are included in the reanalysis post processing for each of the years.

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Urban	1	1	1	1	1	31	40	39	39	40	42	41	40	41	42	42	43
Suburban	0	0	0	0	0	2	2	2	2	2	2	2	3	2	2	2	2
Rural	2	2	2	3	3	3	3	3	3	3	3	3	3	4	3	5	5
Total	3	3	3	4	4	36	45	44	44	45	47	46	46	47	47	49	50

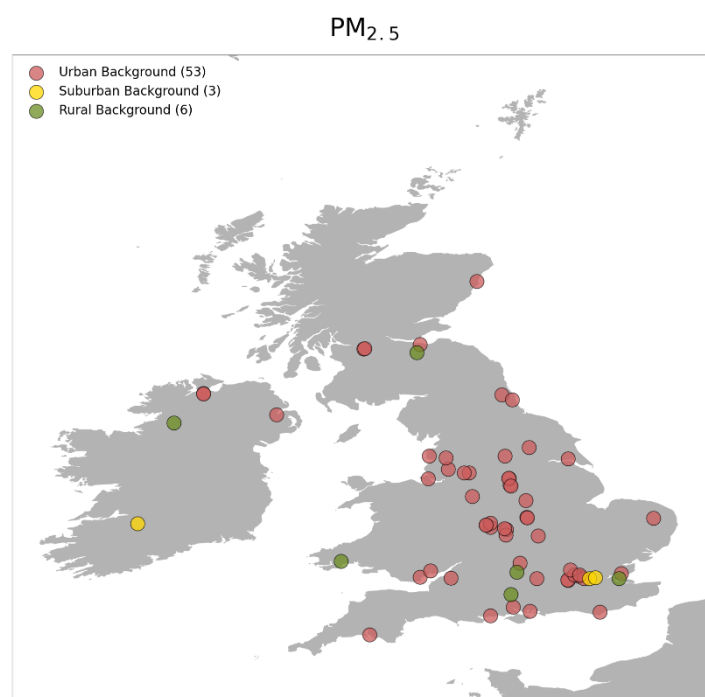


Figure 3.6: Map illustrating the locations of the background sites which measure $PM_{2.5}$ during the current reanalysis period (2003 – 2019).

3.2.7. Coarse Particulate Matter (PM_{10})

Table 3.7: Number of background sites (urban, suburban, rural and total) which measure PM_{10} and are included in the reanalysis post processing for each of the years.

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Urban	45	45	43	43	42	38	35	30	30	30	31	29	29	28	27	26	39
Suburban	3	3	3	3	3	1	0	0	0	0	0	0	1	0	0	0	0
Rural	4	4	4	5	5	5	5	4	5	5	5	5	5	6	5	5	5
Total	52	52	50	51	50	44	40	34	35	35	36	34	35	34	32	31	44

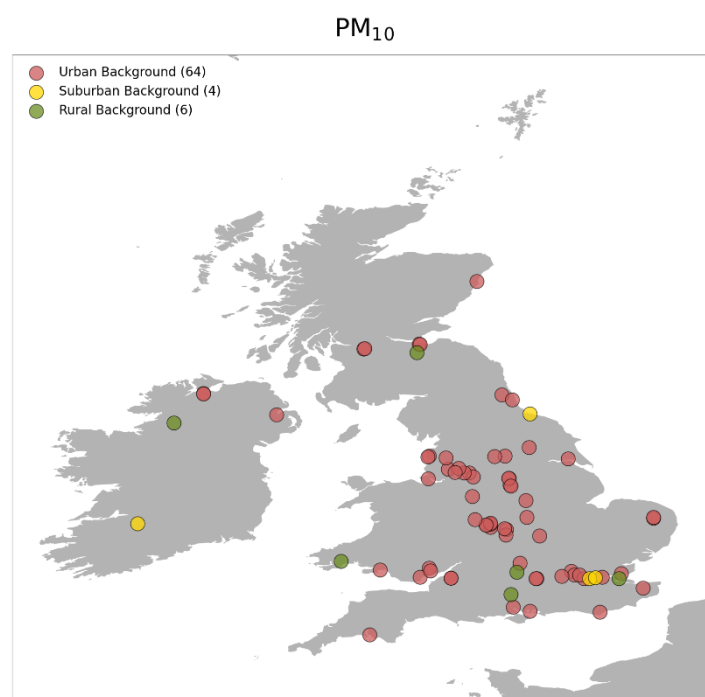


Figure 3.7: Map illustrating the locations of the background sites which measure PM_{10} during the current reanalysis period (2003 – 2019).

4. Dataset Contents

4.1. Dataset Domain

The horizontal domain of the reanalysis data files covers the UK, Republic of Ireland and a section of Northern France in order to capture the Channel Islands. The domain covers an area from -11.0°E to $+2.5^{\circ}\text{E}$ longitude and $+49.0^{\circ}\text{N}$ to $+61.5^{\circ}\text{N}$ latitude. This domain is illustrated in Figure 4.1.



Figure 4.1: Horizontal domain of the reanalysis dataset.

The reanalysis data has been re-gridded from the native model rotated pole coordinate system to a WGS 84 coordinate system with a horizontal resolution of 0.1° degree (approximately 10 km).

4.2. Surface Level Data

The surface level data files cover the whole reanalysis period (1st January 2003 to 31st December 2019) at hourly time resolution for the whole domain detailed above. The time stamps associated with the data are 'hour ending'.

The data is saved into monthly NetCDF files and the file names take the form:

YYYYMM_UK_Air_Quality_Reanalysis_Surface_Level.nc

4.2.1. Data Description

Data Type	Gridded
Horizontal Coverage	West Boundary = -11.0 °E East Boundary = +2.5 °E South Boundary = +49.0 °N North Boundary = +61.5 °N
Horizontal Resolution	0.1 ° x 0.1 ° (10 km x 10 km)
Vertical Coverage	Surface
Temporal Coverage	2003 - 2019
Temporal Resolution	Monthly files containing 1-hourly analyses
File Format	NetCDF

4.2.2. Main Variables

Name	Units
Ozone (O ₃)	µg m ⁻³
Nitrogen Monoxide (NO)	µg m ⁻³
Nitrogen Dioxide (NO ₂)	µg m ⁻³
Nitrogen Oxides (NO _x) as mass of NO ₂	µg m ⁻³
Sulphur Dioxide (SO ₂)	µg m ⁻³
Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Carbon Monoxide (CO)	µg m ⁻³
Ammonia (NH ₃)	µg m ⁻³
Isoprene (C ₅ H ₈)	µg m ⁻³
Total Non-Methane Volatile Organic Compounds (NMVOC)	ppb

Black Carbon in Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Organic Carbon from Fossil Fuels in Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Secondary Organic Aerosol in Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Biomass Burning Aerosol in Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Dust in Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Sulphate in Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Nitrate in Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Ammonium in Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Ammonium Sulphate in Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Ammonium Nitrate in Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Black Carbon in Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Organic Carbon from Fossil Fuels in Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Secondary Organic Aerosol in Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Biomass Burning Aerosol in Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Dust in Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Sulphate in Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Nitrate in Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Ammonium in Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Ammonium Sulphate in Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Ammonium Nitrate in Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Daily Air Quality Index (DAQI)	N/A
Daily Air Quality Index of Ozone	N/A
Daily Air Quality Index of Nitrogen Dioxide	N/A
Daily Air Quality Index of Sulphur Dioxide	N/A
Daily Air Quality Index of Particulate Matter d < 2.5 µm	N/A
Daily Air Quality Index of Particulate Matter d < 10 µm	N/A
X Wind at 10 m	m s ⁻¹
Y Wind at 10 m	m s ⁻¹
Wind Speed at 10 m	m s ⁻¹
Wind Direction at 10 m	degrees
Surface Air Pressure	Pa
1.5 m Air Temperature	K
1.5 m Relative Humidity	%

Surface Upward Sensible Heat Flux	W m ⁻²
Precipitation Amount	kg m ⁻²
Atmosphere Boundary Layer Thickness	m
Very Low Cloud Fraction	N/A
Low Cloud Fraction	N/A
Medium Cloud Fraction	N/A
High Cloud Fraction	N/A

4.3. Bias Corrected Surface Level Data

The bias corrected surface level data files cover the whole reanalysis period (1st January 2003 to 31st December 2019) at hourly time resolution. The time stamps associated with the data are 'hour ending'. The pollutant observations included in the bias correction are only available for land-based locations in the UK and the Republic of Ireland so for these data files, a land-sea mask has been applied to remove data for locations over the ocean. Additionally the data over the area of Northern France that falls within the domain is also removed in these data files.

The data is saved into monthly NetCDF files and the file names take the form:

YYYYMM_UK_Air_Quality_Reanalysis_Bias_Corrected_Surface_Level.nc

4.3.1. Data Description

Data Type	Gridded
Horizontal Coverage	West Boundary = -11.0 °E East Boundary = +2.5 °E South Boundary = +49.0 °N North Boundary = +61.5 °N
Horizontal Resolution	0.1 ° x 0.1 ° (10 km x 10 km)
Vertical Coverage	Surface
Temporal Coverage	2003 - 2019
Temporal Resolution	Monthly files containing 1-hourly analyses
File Format	NetCDF

4.3.2. Main Variables

Name	Units
Ozone (O ₃)	µg m ⁻³
Nitrogen Monoxide (NO)	µg m ⁻³
Nitrogen Dioxide (NO ₂)	µg m ⁻³
Nitrogen Oxides (NO _x) as mass of NO ₂	µg m ⁻³
Sulphur Dioxide (SO ₂)	µg m ⁻³
Particulate Matter d < 2.5 µm (PM _{2.5})	µg m ⁻³
Particulate Matter d < 10 µm (PM ₁₀)	µg m ⁻³
Carbon Monoxide (CO)	µg m ⁻³

Daily Air Quality Index (DAQI)	N/A
Daily Air Quality Index of Ozone	N/A
Daily Air Quality Index of Nitrogen Dioxide	N/A
Daily Air Quality Index of Sulphur Dioxide	N/A
Daily Air Quality Index of Particulate Matter d < 2.5 µm	N/A
Daily Air Quality Index of Particulate Matter d < 10 µm	N/A

5. Acronyms

AQCOPS	Copernicus Europe Domain AQUM
AQEUR	UK and Northwest Europe Domain AQUM
AQREAN	Reanalysis Setup of the UK and Northwest Europe Domain AQUM
AQUM	Air Quality in the Unified Model
AURN	Automatic Urban and Rural Network
C ₅ H ₈	Isoprene
CAMS	Copernicus Atmosphere Monitoring Service
CAMS-GLOB-BIO	CAMS Global Biogenic Inventory
CAMS-REG-AP	CAMS Regional European Anthropogenic Inventory
CLASSIC	Coupled Large-scale Aerosol Simulator for Studies in Climate
CO	Carbon Monoxide
DAQI	Daily Air Quality Index
EAC4	ECMWF 4 th Generation Atmospheric Composition Reanalysis
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	ECMWF 5 th Generation Reanalysis
GFAS	Global Fire Assimilation System
LBC	Lateral Boundary Condition
NetCDF	Network Common Data Form
NH ₃	Ammonia
NMVOC	Non Methane Volatile Organic Compound
NO	Nitric Oxide
NO ₂	Nitrogen Dioxide
NO _x	Nitrogen Oxides
O ₃	Ozone
PM	Particulate Matter
PM ₁₀	Particulate Matter with a diameter < 10 µm
PM _{2.5}	Particulate Matter with a diameter < 2.5 µm
RAQ	Regional Air Quality
SO ₂	Sulphur Dioxide

SPF	Strategic Priorities Fund
SPPO	Statistical Post Processing of Observations
UK	United Kingdom
UKCA	United Kingdom Chemistry and Aerosols
UM	Unified Model
WGS	World Geodetic System

6. References

N. H. Savage, P. Agnew, L. S. Davis, C. Ordóñez, R. Thorpe, C. E. Johnson, F. M. O'Connor and M. Dalvi, Air quality modelling using the Met Office Unified Model (AQUUM OS24-26): model description and initial evaluation, *Geosci. Model Dev.* **2013**, 6, 353-372, <https://doi.org/10.5194/gmd-6-353-2013>

L. S. Neal, P. Agnew, S. Moseley, C. Ordóñez, N. H. Savage and M. Tilbee, Application of a statistical post-processing technique to a gridded, operational, air quality forecast, *Atmospheric Environment*, **2014**, 98, 385-393, <https://doi.org/10.1016/j.atmosenv.2014.09.004>

Carslaw, D. C. and K. Ropkins, openair --- an R package for air quality data analysis. *Environmental Modelling & Software*, **2012**, Volume 27-28, 52-61