ARPHA Preprints

PENSOFT

Research Article

Author-formatted document posted on 26/10/2023

Published in a RIO article collection by decision of the collection editors.

DOI: https://doi.org/10.3897/arphapreprints.e114675

Climate and Air Quality Indices for the European Social Survey

Hilde Orten, David Rayner, Eirik Stavestrand, Iris Alfredsson, Ilze Lace, Irena Vipavc Brvar, Maja Dolinar, Joachim Wackerow, Hannah Clark

Climate and Air Quality Indices for the European Social Survey

Authors: David Rayner, Eirik Stavestrand, Iris Alfredsson, Ilze Lace, Irena Vipavc Brvar, Maja Dolinar, Joachim Wackerow, Hilde Orten

Keywords: Attitudes; Behaviors; Climate; Air quality; Data integration; ESS; ERA5; EEA;

Summary

Knowledge of how personal experience with climate and air quality influence personal attitudes, concerns, and actions about environmental issues is increasingly important. A solid foundation for such studies is to combine interview or survey data on respondents' attitudes and beliefs with indices created from independent meteorological or environmental monitoring data matched to the respondents location.

In this project, indicators of climate and air pollution were integrated with data from the European Social Survey for a selection of large European urban regions. A prototype provenance description application was also developed for describing the workflow for creating indicators and integrating data.

Our main focus was on creating indicators that represent regional anomalies in local air quality and weather for a range of time windows up-to-and-including the dates of the interviews. The goal is to facilitate investigation of relationships between urban citizen's attitudes and behaviors as represented in the survey responses and the conditions in their local environment.

1) Description of the project

Climate Neutral and Smart Cities is the 9th of EOSC Future Task 6.3's Science Project, funded by EU's Horizon 2020 programme[1]. The main objective of the Science Project (SP) is to demonstrate that environmental data and data about people's attitudes, behavior and involvement can be combined for social, political and scientific analysis.
The project rests on three pillars:

1) Data integration and indicator production (the topic covered by this paper).
2) FAIR and structured metadata standards for interdisciplinary use.
3) Dissemination of data and metadata through a dedicated application, available as a service from the EOSC Portal and the EOSC Marketplace.

[1] This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017536

The project involves scientists from the Social Sciences & Humanities Open Cloud (SSHOC) and Environmental Research Infrastructures (ENVRI) working together to produce new and useful outputs for the benefit of the research community, such as indicators of climate and environmental conditions, and methods and workflows for computing them.

Indices for climate and air pollution have been integrated with data from the European Social Survey for a selection of large European urban regions. The data and metadata are currently accessible through the ESS Data Portal[2]. A prototype provenance description application describing the workflow for creating indicators and integrating data was also developed[3].

The data from the project will also become available through a prototype exploratory “labs” service from the European Open Science Cloud (EOSC) Portal and the EOSC Marketplace. This will allow easy access to cross-domain data for scientific analysis and their provenance, as well as to other deliverables from the project.

Partners in the project are the European Social Survey (ESS) ERIC, represented by City University of London and Sikt - Norwegian Agency for Shared Services in Education and Research, Consortium of European Social Science Data Archives (CESSDA), represented by Swedish National Data Service (SND) and the ADP - Slovenian Social Science Data Archives, and the Environmental Research Infrastructures (ENVRI) Community, represented by the In-service Aircraft for a Global Observing System (IAGOS). Experts from The Climate and Environmental Research Institute (NILU) and the Norwegian Meteorological Institute (MET Norway) have provided advice related to the environmental data used in the project.

2) Background

With climate changes caused by human activities affecting weather and climate extremes in many regions (Calvin et al., 2023), knowledge about how personal experience of climate change and extremes could influence personal attitudes, concerns and actions is increasingly important. Recent research has focused on how experience of climate extremes could influence attitudes to environmental issues, especially climate change.

Experiences are important. A summary of the psychological and cognitive basis for how climate experiences might influence beliefs and risk perception is provided by Hoffmann et al. (2022). They reason that direct exposure to climate extremes can make impacts seem more certain, more immediate, and more likely to affect the person themself and their neighborhood rather than an unrelated social group.

Studying how personal experience of extreme weather influences concern for climate change is clearly amenable to interview and survey-based research methodologies. However, using purely self-reporting-based research methodologies may be complicated by *perception bias*: respondents who are more concerned about climate change may also be more likely to perceive that they have experienced extreme events. Leiserowitz et al. (2012) found that respondents who are more concerned about climate change are more likely to report recalling unusual weather events in their local area. Similarly, Whitmarsh (2008) investigated the relationship between experience of air

[2] https://ess-search.nsd.no/en/study/71586b4f-ef66-4b90-aed7-e7e7ad7406ce
[3] https://eosc-provenance.sikt.no/

pollution and climate change concern using the respondent's own evaluation of air pollution's health effects. The study found the respondent's experience of air pollution effects was significantly related to concern about anthropogenic climate change, but with this methodology it is not possible to untangle cause and effect.

A more objective measure of extreme climate experience might be obtained by asking respondents if they have suffered damages due to such events. The above mentioned (Whitmarsh, 2008) study also investigated the relationship between flooding and concern for climate change for a flood-affected region in Hampshire, England. Respondents were classified as flood victims based on if they reported having suffered damage to homes, gardens or vehicles. Although the study found that concern about anthropogenic climate change was significantly related to perception of experience of air pollution effects, there was little relationship with personal experience of flooding. This is not to say there were no differences at all: flood victims were more likely to look at climate information when they came across it in the media; more likely to consider climate change to be personally "very important"; and more likely to expect climate change to affect them through flooding. In general, however, flood victims tended to attribute their flooding experience primarily to local development or maintenance issues, rather than human-induced climate change, at least in the complementary interview studies.

The study of Dai et al. (2015) is interesting in this context because it considered separately respondents who stated they had experience of extreme weather events but suffered no damages, and those who stated they had suffered physical or financial damages as a result. The experienced extreme events included heatwaves, heavy rainfalls or floods, droughts, sandstorms, windstorms, and even avalanches, but damages were mostly reported for heatwaves and heavy rainfalls or floods. The results showed clearly that perceived experiences with extreme weather events were positively related to global climate change beliefs, but with an even stronger relationship for those suffering damages.

But another approach to minimizing perception bias is to combine interview or survey data about respondents' attitudes and beliefs with independent meteorological or environmental monitoring data from the physical sciences.

An early study from the United States (Hamilton & Stampone, 2013) found that, amongst poll respondents who declared themselves independent voters[4], beliefs about climate change were predicted by temperature anomalies for the day and day-prior-to their interview. Brooks et al. (2014), conversely, found a U-shaped response in concern about climate change to temperature anomalies on the day-of-interview, with higher concern both for respondents experiencing anomalously warm and anomalously cold temperatures. The apparent discrepancy is perhaps explained by the inclusion of much more extreme anomalies (~ ± 15 °C) in the Brooks et al. data than the Hamilton & Stampone study (~ ±6 °C). Bergquist & Warshaw (2019) found that even state-wide annual temperature anomalies in the US were associated with climate concern as expressed in poll results, with a 1 °C increase in temperature corresponding to an increase in "about 1% in people who worry a *great deal or fair amount* about climate change". They did not see any comparable correspondence with standardized indicators of precipitation extreme events, however.

More recently, Hoffmann et al. (2022) presented a Europe-wide study on the impact of climate extremes on environmental concern and Green voting. They constructed climate indices based on: mean temperature; the Universal Thermal Comfort Index, which represents the human physiological

[4] not aligned with a political party

response to temperature, humidity, wind and solar radiation; and the Standardized Precipitation-Evapotranspiration Index (SPEI), representing water balance. The indices used were: monthly mean temperature anomaly; heat episodes, based on the number of days where at least three consecutive days with mean temperature or UTCI were above the local monthly 95th percentile; and dry/wet spells, for which monthly mean SPEI < -0.5/ SPEI >+0.5. All indices were relative to a 1971-2000 reference period. Hoffman et al. found significant impacts but regional differences across Europe. In cold and temperature regions, the strongest impacts were from positive temperature anomalies, heat episodes and dry spells, whereas in hot regions the relationships were generally not statistically-significant. They also investigated what time-window would maximize these relationships and found higher relationships for a 12-month window than for 1, 6, or 24 month windows. Their interpretation was that a single month of anomalous weather is not sufficient to increase environmental concerns, and that an extended period is required. As events become more distant in time, however, their impacts become less, leading to an "inverted U" shaped response curve.

3) Data sources

The data used for the project come from recognized and open sources within the environmental and social sciences: The European Social Survey (ESS), the European Environmental Agency (EEA) and the ECMWF Reanalysis v5 (ERA5) from the Copernicus program. The ESS was the starting point for the project, whereas EEA and ERA5 were chosen as data sources due to their recognition and uptake, the availability of relevant data covering the required time periods and geographic regions, and the provision of Application Program Interfaces (APIs) allowing programmable data download.

a) ESS Social Survey data

The European Social Survey (ESS)[5] is a biannual social attitude survey covering more than thirty European countries. As part of the SSHOC consortium, ESS collects data through face-to-face interviews[6] with representative samples of respondents from European countrieson topics directly related to the mission, such as citizens' involvement and democracy, political values and engagement, health and social care, and the smart economy and smart mobility. ESS covers political and social trust, health and health inequality, attitudes towards climate change and energy, understanding and valuing democracy, and digital communication at work and in the family, among many other topics. Interviews are usually conducted between September and January, but there are exceptions.

The 8th round of the ESS data data, released 2016, fielded a rotating questionnaire module about attitudes to climate change and energy-use[7]. Data from this module will be particularly interesting for analyses together with data about the physical environment in which the respondents live. The ESS data used in this project therefore starts with the ESS 8th round, 2016 data, and covers all rounds and topics fielded thereafter.

[5] See https://ess-search.nsd.no/
[6] In ESS Round 10 part of the interviews were fielded as self-completion, due to the COVID situation
[7] See section D in the ESS 8 2016 questionnaire at https://stessrelpubprodwe.blob.core.windows.net/data/round8/fieldwork/source/ESS8_source_questionnaires.pdf and the background in https://www.europeansocialsurvey.org/docs/round8/questionnaire/ESS8_climate_final_module_template.pdf

In order to match individual survey responses with air quality or climate data, survey responses must be tagged with temporal and spatial information. Data from ESS are tagged with geographic information in the form of Nomenclature of territories for statistics (NUTS)[8] codes, and the time and date of the interview. Further details of the NUTS regions used are provided in Section 4, Study regions.

b) EEA air quality data

Motivated by the well-demonstrated effects of air pollution on health (Schraufnagel et al., 2019), we aim to provide data that allow examining the effects of local air quality on well-being and perception of one's health and attitudes toward other issues.

The European Environmental Agency (EEA) European Air Quality Index (AQI)[9] categorizes the concentrations of pollutants in terms of associated health impacts (van den Elshou et al., 2012). This index is thus ideal for the project, but Air Quality Index data are currently only provided by EEA as real-time data. However, the EEA provides extensive access to historical air pollutant concentration data[10] . Thus, we downloaded historical EEA air pollution data for the regions and time-periods matching the ESS data and calculated the AQI and its components ourselves following the EEA documentation (cite same as above). The pollutants used to calculate the AQI are shown in Table 1.

Data from in situ background stations in the selected cities were used. Only background stations were selected because they were assumed to provide better measures for the overall air quality in an area than, for example, traffic stations located near roads, and industrial stations located near industrial areas. Hourly data from all background stations - including urban, suburban and rural stations - within the NUTS region extents were downloaded.

Table 1: the pollutants for which concentration measurements were used to calculate air quality indicators.

Pollutant ID	**Pollutant name**	**Units**
PM_{10}	Particulate matter	µg/m^3
$PM_{2,5}$	Fine particulate matter	µg/m^3
NO_2	Nitrogen Dioxide	µg/m^3
SO_2	Sulfur Dioxide	µg/m^3
O_3	Ozone (ground - level)	µg/m^3

c) ERA5 climate data

The inclusion of questions about attitudes to climate change in the ESS allows investigation of whether personal experience of climate change and weather extremes could influence responses.

[8] See https://ec.europa.eu/eurostat/web/nuts/background
[9] https://airindex.eea.europa.eu/Map/AQI/Viewer/#
[10] https://discomap.eea.europa.eu/map/fme/AirQualityExport.htm

Our aim was to calculate indicators that represent the climatic experiences of survey respondents prior to their interviews.

Our indicators are based on data from the European Centre for Medium Range Weather Forecasts (ECMWF) reanalysis (ERA5, Hersbach et al.,(2020)). Reanalyses are widely used in climate studies and combine sparse and irregular meteorological observations with a forecast model to provide a best-estimate of atmospheric variables. The ERA5 global reanalysis gridded dataset has ~30km horizontal resolution and covers the period 1940 to present.

As explained in Section 5, in this project we focus on climate indicators derived from temperature, precipitation and wind gust (see Table 2). Gridded single-level data[11] covering the NUTS region extents with a 0.05° buffer were downloaded at 0.1°x 0.1° spatial resolution using the Climate Data Store (CDS) Application Program Interface (API)[12].

Table 2: The ERA5 variables used to construct indicators. More detailed descriptions are available from the ERA5 website.

Climate related variables		
Name	**Label**	**Description**
t2m	2 metre temperature	This variable describes the temperature of air at 2m above the surface of land, sea or in-land waters.
tp	Total precipitation	Precipitation accumulations are over the hour (the processing period) ending at the validity date/time.
i10fg	Instantaneous 10 meter wind gust	This parameter is the maximum wind gust at the specified time, at a height of ten meters above the surface of the Earth.

4) Study regions

The first task was to select cities of interest to include in the integrated data for our project. The following selection criteria were defined: Cities should be large European cities, located in different areas of Europe, and have enough respondents in the ESS to allow statistically-significant relationships to be identified in the unstratified survey responses at a regional level[13].

We then needed to define the regional granularity, and as data from the European Social Survey is available on the Nomenclature of territories for statistics (NUTS)[14] levels 1 to 3, we decided to merge data from the three sources on the lowest NUTS level possible for each of the chosen cities. For some of the selected cities the number of ESS respondents in the NUTS level 3 region would permit

[11] https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview
[12] https://cds.climate.copernicus.eu/api-how-to
[13] While the number of respondents in the cities can be regarded as adequate, we anticipate that analyses using stratification by age/gender/education etc. may represent a challenge.
[14] See https://ec.europa.eu/eurostat/web/nuts/background

reasonable analyses (Table 3). For cities with less ESS respondents available per city NUTS level 2 or 3 needed to be chosen.

Table 3: The selected cities, the NUTS levels used in the project, as well as the number of ESS respondents in each region in 2016.

City	NUTS 2016 region	NUTS level	Number of ESS respondents 2016
Stockholm	SE110/SE11 Stockholms län	3/2[15]	246
Berlin	DE3 Berlin	1	125
Praha	CZ010 Hlavní město Praha	3	277
Budapest	HU110 Budapest	3	291
Wien	AT13 Wien	2	398
Madrid	ES30 Comunidad de Madrid	2	167
Paris	FR10 Île de France	2	158
Bruxelles/ Brussels	BE10 Région de Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest	2	182
London	UKI London	1	143
Oslo	NO01 Oslo and Akershus[16]	2	251

For the purpose of the project, the 2016 version of NUTS was used. This is because ESS data used in this project included data from 2016 to 2022, and both the preceding NUTS version 2013, and the successor version of NUTS 2021, contain only minor differences from the NUTS 2016 nomenclature for the selected cities[17]. For Hungary, for example, the Code HU101 in NUTS 2013 changed to HU110 in NUTS 2016 with no associated territorial changes.

For Norway, which is an EFTA country, an extensional regional reform took place in 2020. At NUTS level 2 'Oslo og Akershus' in NUTS 2016 changed to 'Oslo and Viken' in NUTS 2021, with the latter representing a much larger area. ESS data from Round 10 (2020 – 2022) are collected using the 2021 version of NUTS, and is regarded as not comparable and is therefore not included in the integrated data set from our project for ESS round 10.

[15] Sweden provided data at NUTS level 3 for ESS round 8 and NUTS level 2 for ESS round 9 and 10. For Stockholm the geographical polygons are, however, the same.

[16] While the NUTS classification is defined only for the Member States of the EU, Eurostat, in agreement with the countries concerned, also defines a coding of statistical regions for countries that do not belong to the EU but are either, candidate countries awaiting accession to accession to the EU, potential candidates or countries belonging to the European Free Trade Association (EFTA).

[17] See https://ec.europa.eu/eurostat/web/nuts/history

All climate and air pollution indices are provided as regional values in the integrated datasets. The NUTS 2016 polygons available from the related Eurostat GISCO[18] pages have been used to define the selected city regions.

5) Index specifications

a) Temporal considerations

The goal of the project is to make it easier for researchers to explore relationships between urban citizen's attitudes and behaviors as represented in the survey responses and the local environment in which they live. Our main focus has been on creating indicators that represent regional anomalies in local air quality and weather for a range of time windows up-to-and-including the dates of the interviews. As discussed above, the temporal aspect of how experience of air pollution and weather anomalies impacts attitudes is far from well-defined. However, we subscribe to the basic reasoning from Hoffmann et al. (2022), that an extended period of anomalous weather is probably required to increase environmental concern, but that as events become more distant in the past, their impact recedes. Our approach has thus been to characterize regional conditions in specific time-slots relative to the dates of the interviews. Our primary climate indices represent conditions for every date, the week (7 days) before the date, a month (30 days) before the date, three months (90 days) before the date, and year (365 days) before the date. Our primary air pollution indices represent conditions for every date, the 2 days including the interview, the week (7 days) before the date, a month (30 days) before the date, and year (365 days) before the date.

b) Climate data indices

Our choice of regional climate indicators are intended to capture both seasonal and shorter-duration weather anomalies using measures that are easily understood and easily interpreted. We decided to base our climate indices on temperature, precipitation and wind gust (as a proxy for storminess) variables so they can be immediately understood by interdisciplinary researchers. The indices are not intended to indicate whether specific weather-related extreme events (e.g. heat waves, cloudbursts) have occurred during the time-window. The full list of climate indices in the dataset is provided in Table 4.

The climate indices are based on daily regional averages (or maximums, for wind gust indices) . Note that when aggregating hourly ERA5 data to daily values, all hourly measurements from the day have been included, including those from after times when interviews on that date took place.

Our primary indices are regional average temperature, total precipitation and instantaneous 10 meter maximum wind gust for time-windows leading up to the date of each interview. For temperature, indices based on daily minimums and maximums have also been created. Changes in average temperature are the most immediate and direct effect of anthropogenic climate change, and recent time-periods with anomalous average temperature or precipitation are generally identifiable from personal experience. Anomalies in average temperature or total precipitation for the longer time-windows (3-month or year) may not, however, reflect anomalous weather

[18] See https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts

sequences that include both colder-than-average and warmer-than-average conditions, or both dry and wet conditions.

The climate averages obviously vary dramatically across the cities in the study, and the sub-yearly indices also have seasonal fluctuations. Studies that involve intercomparison between the cities in the study will require more standardized indices.

As standardized measures of temperature variability, we provide temperature anomalies for the date of interview and month of interview, calculated with-respect-to the multi-year calendar month average for the climate baseline period 1991 -2020. In addition, we provide the calendar-month 95th percentile temperature for the baseline 1991 - 2020, with which the other temperature measures can be compared.

As a measure of standardized precipitation, we provide the relative total precipitation, calculated as the precipitation for the calendar month of the interview divided by the multi-year calendar month average for the 1991 - 2020 baseline period, multiplied by 100.

No standardized wind gust anomaly indices have been created by us. On the other hand, the available baseline reference variable will allow researchers to do so as they find useful.

Graphical representations of selected climate indices are presented in Figures A1 – A3 in Appendix A.

Table 4: The full list of climate indices in the dataset[19].

Index	Index Label
tmpdca	Temperature in degrees Celsius, date average
tmpdcmx	Temperature in degrees Celsius, date maximum
tmpdcmn	Temperature in degrees Celsius, date minimum
tmpdcaw	Temperature in degrees Celsius, week average before the date
tmpdcam	Temperature in degrees Celsius, month average before the date
tmpdca3m	Temperature in degrees Celsius, three months average before the date
tmpdcay	Temperature in degrees Celsius, year average before the date
tmpdcacm	Temperature in degrees Celsius, calendar month average
tmpdcamb	Temperature in degrees Celsius, multi-year calendar month averages, baseline 1991 - 2020
tmp95pacmb	Temperature in degrees Celsius, multi-year calendar month 95th percentiles, baseline 1991 - 2020
tmpanod	Temperature anomaly date
tmpanocm	Temperature anomaly calendar month

[19] The links leads to a description of the process step related to the creation of the variable in the DDI-CDI Process description tool (see section 8)

paccta	Total precipitation average, date
pacctaw	Total precipitation, weekly sum to date.
pacctam	Total precipitation, monthly sum to date.
paccta3m	Total precipitation, three-monthly sum to date.
pacctay	Total precipitation, yearly sum to date.
pacctcm	Total precipitation, calendar month
pacctmb	Total precipitation, multi-year calendar month averages, baseline 1991 - 2020
paccdcm	Total precipitation, calendar month relative to normal.
iwg10mx	Instantaneous 10 metre maximum wind gust, date
iwg10mxaw	Instantaneous 10 metre maximum wind gust maximum, week average to date.
iwg10mxam	Instantaneous 10 metre maximum wind gust maximum, month average to date.
iwg10mxa3m	Instantaneous 10 metre maximum wind gust maximum, 3-month average to date.
iwg10mxay	Instantaneous 10 metre maximum wind gust maximum, year average to date.
iwg10mxamb	Instantaneous 10 metre maximum wind gust maximum, multi-year calendar month averages, baseline 1991 - 2020

c) Air quality indices

For the air quality data we start by creating regional European Air Quality Index[20] (AQI) components for each of the identified pollutants (PM_{10}, $PM_{2.5}$, SO_2, NO_2 and O_2). The worst air quality level by date and region for each pollutant was calculated from the hourly concentration measurements using the 99th percentile of concentrations for each date. The index level for each hour and pollutant was calculated according to Table 5. Note that the classification includes a maximum pollutant concentration above which the measurement is assumed to be erroneous, but no values exceeding these maximums were present in the data used in this project. Graphical representations of these time-series are provided in Figures A4 – A8 in Appendix A. Then the regional Air Quality Index itself was calculated as the worst index for the date over all the pollutant-specific indices. The procedure was repeated to calculate the AQI for the last two days.

[20] See https://airindex.eea.europa.eu/Map/AQI/Viewer/. Categories are 'Good', 'Fair', 'Moderate', 'Poor', 'Very poor' and 'Extremely poor'

Table 5: The European Air Quality Index component classifications for the individual pollutants. The numeric values for each category are the values used in our datasets.

Pollutant	Index level (based on pollutant concentrations in µg/m3)					
	0 = Good	1 = Fair	2 = Moderate	3 = Poor	4 = Very poor	5 = Extremely poor
PM_{10}	0-20	20-40	40-50	50-100	100-150	150-1200
$PM_{2,5}$	0-10	10-20	20-25	25-50	50-75	75-800
NO_2	0-40	40-90	90-120	120-230	230-340	340-1000
SO_2	0-100	100-200	200-350	350-500	500-750	750-1250
O_3	0-50	50-100	100-130	130-240	240-380	380-800

Because the AQI represents the worst index observed during a short time-period, it is not suitable to extend the direct calculation of AQI beyond two days. For air quality indices for long time-windows, we instead calculated the number of days with poor-or-worse air quality in the region by pollutant and across pollutants, for each week (7 days), month (30 days) and year (365 days) prior to every date. The full list of air quality indices in the dataset is given in Table AQV. The rationale for including variables for each of the individual pollutants in the dataset is to allow identifying which pollutants contribute to the aggregated Air Quality Index; these variables are not primarily intended for comparison with survey response variables.

Table 6: The full list of air quality indices in the dataset[21].

Index	**Index Label**
aqiwdpm10	Worst air quality index level PM10, date
aqiwdpm2_5	Worst air quality index level PM2.5, date
aqiwdso2	Worst air quality index level SO2, date
aqiwdno2	Worst air quality index level NO2, date
aqiwdo3	Worst air quality index level O3, date
aqiwd	Air Quality Index (AQI). Worst air quality index level across pollutants, date
aqiw2dpm10	Worst air quality index level PM10, last two days
aqiw2dpm2_5	Worst air quality index level PM2.5, last two days
aqiw2dso2	Worst air quality index level SO2, last two days
aqiw2dno2	Worst air quality index level NO2, last two days
aqiw2do3	Worst air quality index level O3, last two days

[21] The links leads to a description of the process step related to the creation of the variable in the DDI-CDI Process description tool (see section 8).

aqiw2d	Worst air quality index level across pollutants, last two days
ndyprwpm10	Number of days with 'poor' air quality level or worse on PM10, week before the date
ndyprwpm2_5	Number of days with 'poor' air quality level or worse on PM2.5, week before the date
ndyprwso2	Number of days with 'poor' air quality level or worse on SO2, week before the date
ndyprwno2	Number of days with 'poor' air quality level or worse on NO2, week before the date
ndyprwo3	Number of days with 'poor' air quality level or worse on O3, week before the date
ndyprw	Number of days with 'Poor' air quality level or worse on one or more pollutant indicators, week before the date
ndyprmpm10	Number of days with 'poor' air quality level or worse on PM10, month before the date
ndyprmpm2_5	Number of days with 'poor' air quality level or worse PM2.5, month before the date
ndyprmso2	Number of days with 'poor' air quality level or worse SO2, month before the date
ndyprmno2	Number of days with 'poor' air quality level or worse NO2, month before the date
ndyprmo3	Number of days with 'poor' air quality level or worse O3, month before the date
ndyprm	Number of days with days with 'poor' level or worse on one or more pollutant indicators, month before the date
ndyprypm10	Number of days with 'poor' air quality level or worse on PM10, year before the date
ndyprypm2_5	Number of days with 'poor' air quality level or worse on PM2.5, year before the date
ndypryso2	Number of days with 'poor' air quality level or worse on SO2, year before the date
ndypryno2	Number of days with 'poor' air quality level or worse on NO2, year before the date
ndypryo3	Number of days with 'poor' air quality level or worse on O3, year before the date
ndypry	Number of days with days with 'poor' level or worse on one or more pollutant indicators, year before the date

6) Dataset integration

Integrating the climate and air quality indices with data from the ESS, providing data matched to the ESS round, region, and the date of each interview, will allow researchers to explore possible impacts of air quality and climate events during specific time windows relative to the timing of the interview, on the survey responses.

To integrate data from the ESS with gridded ERA5 climate data and EEA air quality data, both geographical and temporal aspects need to be aligned.

a) Geographic aggregation

In order to match data from the three data sources geographically, a common map projection needs to be chosen. As the projection EPSG:4326[22] can be used for all of the datasets, we considered this to be the best choice. EPSG:4326 is the World Geodetic System (WGS)[23]1984 used in GPS. It is an ellipsoidal two dimensional Cartesian coordinate system with latitude/longitude axes, orientation north, east, with degree as the unit of measure. The NUTS 2016 polygons from Eurostat GISCO[24], the ERA5, and the EEA station location information, were all already in EPSG:4326.

Geographic averaging of ERA5 climate data.

The ERA5 gridded data were downloaded on a 0.1 x 0.1 degree latitude/longitude grid. Multiple grid cells covering each of the selected regional polygons were downloaded. Cells on the border of a region may be partly interior and partly exterior to the polygon. Also, within a region, the population is unevenly distributed, such that some of the areas within the region are more densely populated than other parts. For these reasons, weighted averages for the climate variables were calculated based on how much of a grid cell lies within a polygon, as well as the population of each grid cell. The population size within the study region for each grid cell was calculated using population data from Global Human Settlement dataset GHS-POP[25]. This means temperature, precipitation, and wind measures for a grid cell with a larger population will count more towards the regional average than does a cell with a smaller population. An example showing the cell weighting for the region AT13 (Vienna/Wien) is shown in Figure 1.

[22] See https://epsg.io/4326
[23] See https://gisgeography.com/wgs84-world-geodetic-system/

[24] See https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts
[25] https://ghsl.jrc.ec.europa.eu/datasets.php

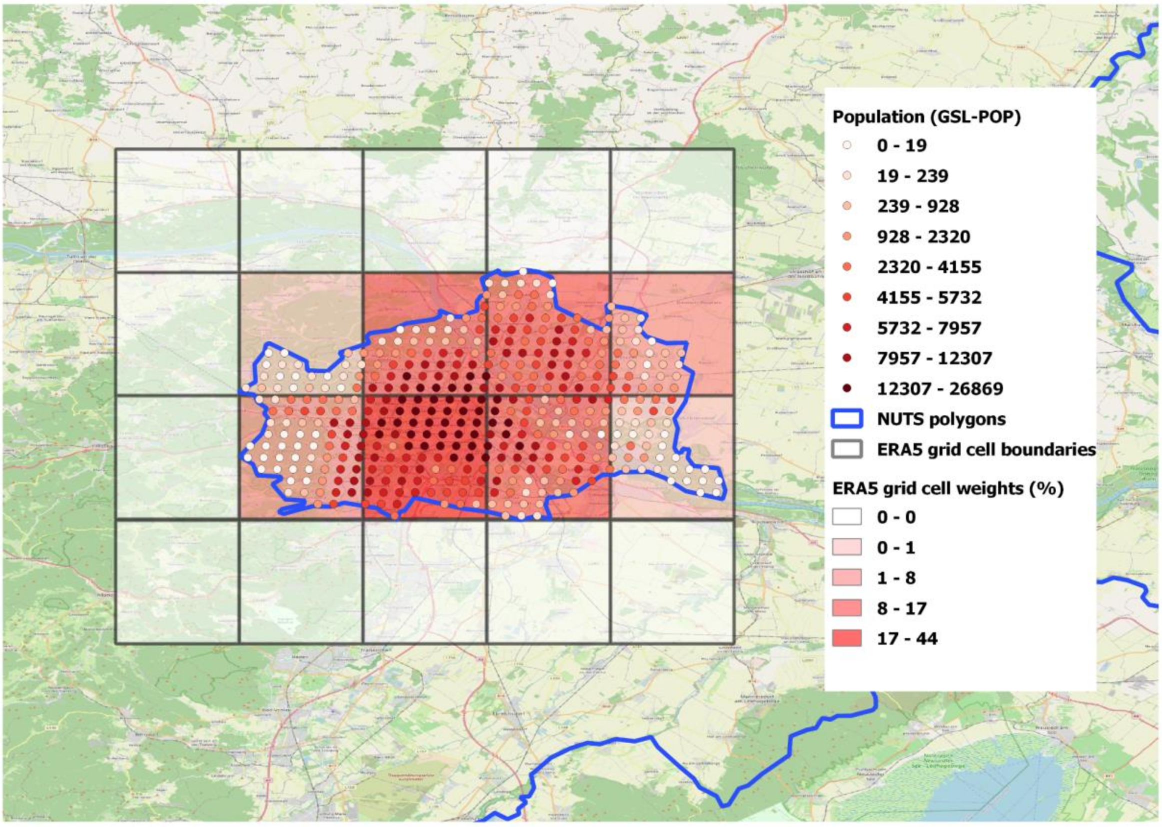

Figure 1: Climate data gridding and weighting example for NUTS region AT13 (Vienna/Wien), Austria. The ERA5 grid cell boundaries show the location and extent of the 0.1°x 0.1° climate data downloaded for the region; colored dots show population density from the GHSL Data Package (GSL-POP) within the NUTS region; colored grid cells show the weightings used for the daily climate data, which are proportional to the average population density in the region within each cell. Note that the downloaded ERA5 data includes a buffer region, which explains the presence of cells in the downloaded data that do not contribute to the regional average.

A similar approach is used by De Schrijver et al. (2021) for performing epidemiological assessment of the health impact of temperature and climate change. They observed that population-weighted climate variable averages derived from high resolution gridded climate datasets (ERA5 data also in this case) provided a better approximation of the true exposure of the population compared to unweighted data, especially in densely populated urban areas with large intra-city temperature variability.

As our project is aiming at providing data that facilitates the study effects of climate on citizens attitudes and behaviors, we consider the approach of using gridded population weighted climate datasets to be the best approach. If our goal was merely to analyze variations in climate across city regions without combining climate data with data about people, an unweighted approach would, however, have been preferred.

Geographic averaging of EEA air quality data

The air quality data from the EEA are from measurement stations located at specific geographical locations. All background stations that were geographically located within the polygons of the NUTS regions were chosen.

Population weighting of air quality data from the EEA turned out to be a more complex proposition than for ERA5 data. Whereas ERA5 grids can be assigned fixed weights, measurement data from observing stations inevitably include missing data and recording breaks. Thus, any population-distance-based weightings would need to be re-calculated on a daily basis. Combined with the high uncertainty in choosing appropriate distance weighting functions for the various pollutants, we therefore decided that the complexity did not justify implementing population weighting. Thus, the air quality data remain unweighted, and the regional average air pollution concentrations were calculated using the unweighted average of all available hourly measurements.

The different treatments of the climate and air quality data may introduce issues when interpreting analysis results. In the future, the best solution may be to utilize grid-based pollutant concentration values from historical air pollution models.

Geographic considerations with ESS survey data

Data from the European Social Survey are available in NUTS, levels 1 to 3. The current version of NUTS is always used. Because data from the ESS only are publicly available on the NUTS level, a more exact location of the dwelling area of each respondent is not possible. This makes it impossible to determine more exactly which physical conditions a respondent has been exposed to. And even if a more granular location of the respondents dwelling would be possible, this would be an approximation to the respondent’s exposure, as long as the movement patterns of the respondent remains unknown.

Eurostat [26] introduces the idea of grid statistics as an alternative to population statistics for administrative areas. Population grids are a powerful tool to describe our society and to study the interrelationships between human activities and the environment. They are particularly useful for analyzing phenomena, and their causes, which are independent of administrative boundaries, similar to air quality and climate data. Grids would form an additional layer which could be mapped to the current administrative boundaries which can change over time. An idea for the ESS is to consider something similar, and make data available on a more granular level, for example on the level of an 0.1 x 0.1 degree cell grid. This would, however, open up other issues related to disclosure risks of directly and indirectly identifiable respondent data that would need to be addressed.

b) Temporal aggregation considerations.

As discussed above, the study time-period was selected to be 2016 to 2022 because the ESS data from round 8 in 2016 contains a questionnaire module related to attitudes to climate change and energy use. Data for 2016 to 2022 were downloaded for all three data sources. In addition, climate data for 1991 to 2020 were downloaded for ERA5 for normalization purposes.

[26] See https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_grids

The European Social Survey records the exact time of the interview down to the level of the second of the start and end of the interview, from which we created the interview date variable. This variable has been used to merge in data from climate from ERA5 and EEA, that are available on an hourly basis, but have been aggregated to the level of the day for integration purposes. This will allow researchers to be able to explore statistical relationships between environmental indicators and interview responses, relative to the date when the interview took place, in a detailed manner. A source of error that may occur when comparing interview responses with values on environmental indicators on the day of the interview is that the values on environmental indicators may change throughout the day. This means that if an interview was fielded early in the day, the respondent will not yet have experienced conditions that have occurred later in the day but that may have influenced the aggregated daily measures.

Local time zones have been chosen in order to reflect the actual start and end of each date in the area where the respondents live. For all selected cities apart from London, the local time zone is Central European Time (CET)[27] (Coordinated Universal Time (UTC)[28] + 1:00) for standard time and Central European Summertime (CEST)[29](UTC + 2:00). London has Greenwich Mean Time (GMT)[30] (UTC + 00:00) during standard time and British Summer Time (BST)[31] (UTC + 1:00). As ERA5 uses UTC, ERA5 data for each region had to be converted to local time zones for the purpose of data integration.

c) The integrated data set

Integrated data resulting from the project will allow researchers to analyze data from the ESS together with the created air quality and climate data indicator variables created through the project. The main output from the project is thus a set of merged datasets, where ESS data at the level of individual responses from rounds 8, 9 and 10 has been complemented with the climate and air pollution index data for the date and location.

Our overall goal with the integrated dataset is to provide a set of pre-calculated climate and air pollution indices that have meaningful relationships with ESS response data so as to lower the threshold for initiating cross-domain studies. Relevant indices require temporal aggregation of data from a time-window prior to the interview date. Calculating new climate and air pollution indices from the daily regional datasets we provide is of course possible, but using the merged datasets with the pre-calculated climate and air pollution indices matched to each interview response will be far simpler.

The integrated data is essentially complete in terms of the air pollution and climate indices, with the exception of SO2 and minor omissions for PM2.5. The extent of missing data is shown in Table 7.

[27] https://en.wikipedia.org/wiki/Central_European_Time
[28] https://en.wikipedia.org/wiki/Coordinated_Universal_Time
[29] https://en.wikipedia.org/wiki/Central_European_Summer_Time
[30] https://en.wikipedia.org/wiki/Greenwich_Mean_Time
[31] https://en.wikipedia.org/wiki/British_Summer_Time

Table 7. The completeness of the integrated dataset.

ESS data set (see Section 8)	Pollutant	Indices affected	Missing data (%)	Regions affected
EOSC merged-EOSC-ESS10	SO_2	aqiwdso2	0.3%	FR10
EOSC merged-EOSC-ESS10SC	SO_2	aqiwdso2, aqiw2dso2	55.4%	DE3, SE11
EOSC merged-EOSC-ESS9e03_1 (ESS9)	SO_2	aqiwdso2, aqiw2dso2	12.6%	NO01, SE11
	NO_2	aqiwdno2	0.4%	NO01, SE11
EOSC merged-EOSC-ESS8e02_2 (ESS8)	SO_2	aqiwdso2, aqiw2dso2	10.8%	HU101, SE110
	$PM_{2.5}$	aqiwdpm2_5, aqiw2dpm2_5	11.0%	HU101, SE110

8) Workflows and implementation

a) Workflow description

The workflow starting from accessing the input data from the EEA and ERA5 data sources to the final integrated ESS/EEA/ERA5 files has gone through a set of activities and steps to reach the final outcome. The complete workflow can be viewed in the DDI-CDI process description tool[32]. The tool, which is an additional deliverable from the project, shows each of the process activities related to the workflow. The *Data Processing* activity contains steps for creating each of the created indicator variables. For each step a user can view its inputs and outputs, and have access to a human readable description reflecting how the variable has been computed. It also provides a link to the source code of the program that was used to create the variable.

In an interdisciplinary project like *Climate Neutral and Smart Cities* we feel that detailed provenance information is particularly important. We see this tool and the underlying standardized process description as a very valuable resource for verifying how data has been created, allowing researchers to understand the provenance of the data and to evaluate the methods and procedures used.

A paper descri(Sikt - Norwegian Agency for Shared Services in Education and Research, 2023e)*Neutral and Smart Cities* Science Project is that in order to succeed on an interdisciplinary

[32] https://eosc-provenance.sikt.no/

project, collaboration between experts from different domains is crucial for bringing in the knowledge required to do the job. EOSC Future has provided funding for our science project allowing this to happen. This has allowed us to develop products we hope that the research community will find useful as well as to establish networks for future collaborations.

b) ESS Data Portal

Data resulting from the project are published in the ESS Data Portal (Sikt, 20213). From the portal datasets from the project can be downloaded and detailed metadata and graphic visualizations related to the variables can be explored.

Metadata in the ESS Data Portal are structured in the DDI-Lifecycle metadata standard[33] and provide a good overview over the content of the available datasets and the variables included.

Section d) below gives an overview of the datasets available from the project.

c) Verification

The code to download the ERA5 and EEA data has been verified by comparison with an independent partial-implementation in Matlab using private climate data analysis libraries. With the exception of differences resulting from the different mathematical definitions of a percentile in Python's numpy package and Matlab, the two implementations generated the same results.

d) Data and Code Availability

The outcome of the process has resulted in the following data sets available from the ESS Data Portal (Sikt, 20213).

EEA hourly station measurements (EOSC eea-stations), Sikt (2023b):
Contains hourly stations air quality data on station level, for background stations from the EEA in the relevant time period and study regions. This file does not contain any of our computed indicator variables. Due to its size, this file is available in Parquet format only, available for download from the Documents section of the project page in the ESS Data Portal.

Regional Air Quality Indices (EOSC eea-regions), Sikt (2023a):
This dataset contains the computed AQI variables on a daily time-step at the regional level. The data is downloadable in SPSS(.sav), DTA and CSV format.

ERA5 hourly gridded climate data (EOSC era5-grids), Sikt (2023c):
Contains hourly climate variables from ERA5 for all extracted grid cells, for the relevant time period and study regions. Due to its size, this file is available in Parquet format only, available for download

[33] https://ddialliance.org/Specification/DDI-Lifecycle/

from the Documents section of the project page in the ESS Data Portal. This file does not contain any of our computed indicator variables.

Regional Climate Indices (EOSC era5-regions), Sikt (2023d):
This dataset contains the computed climate indicator variables on a daily time-step at the regional level. Datafiles are available in SPSS(.sav), DTA and CSV format.

ESS data with Climate and Air Quality Indices (merged):
The indicator data have been merged with ESS data from round 8, 9 and 10 resulting in the following datasets - all available in SPSS(.sav), DTA and CSV formats:
EOSC merged-EOSC-ESS10 (ESS10 face-to-face interview), Sikt (2023h)
EOSC merged-EOSC-ESS10SC (ESS10 self completion), Sikt (2023i)
EOSC merged-EOSC-ESS9e03_1 (ESS9), Sikt (2023g)
EOSC merged-EOSC-ESS8e02_2 (ESS8), Sikt (2023f)

Code availability
The code used to create the data is available from:
https://github.com/sikt-no/ess-labs-data-sp9

The source code of the DDI-CDI Provenance description tool can be found at:
https://github.com/sikt-no/ddi-cdi_process2web

9) Funding

The EOSC Future Science Project *Climate Neutral and Smart Cities* is funded under the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017536.

10) Acknowledgments

Many thanks to people working on the *Climate Neutral and Smart Cities* for input and support:
Eric Harrison (ESS), Hannah Clark (IAGOS), Bodil Agasøster, Benjamin Beuster, Archana Bidargaddi, Arofan Gregory, Hanna Thome Grieg, Carl-Erik Herheim, Knut Kalgraff Skjåk, Joachim Wackerow, Bjarne Øymyr (Sikt)

Many thanks to Hans Olav Hygen, Øystein Godøy and Lara Ferighi (Met.no) and Kjetil Tørseth, Miha Markelj, Ann-Britt Kårstad og Sverre Solhaug (NILU) for their expert advice regarding air quality and climate related data and methods.

11) References

Bergquist, P., & Warshaw, C. (2019). Does Global Warming Increase Public Concern about Climate Change? *The Journal of Politics*, *81*(2), 686–691. https://doi.org/10.1086/701766

Brooks, J., Oxley, D., Vedlitz, A., Zahran, S., & Lindsey, C. (2014). Abnormal Daily Temperature and Concern about Climate Change Across the United States. *Review of Policy Research*, *31*(3), 199–217. https://doi.org/10.1111/ropr.12067

Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P. W., Trisos, C., Romero, J., Aldunce, P., Barrett, K., Blanco, G., Cheung, W. W. L., Connors, S., Denton, F., Diongue-Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Jones, C., … Péan, C. (2023). *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* (First). Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647

Dai, J., Kesternich, M., Löschel, A., & Ziegler, A. (2015). Extreme weather experiences and climate change beliefs in China: An econometric analysis. *Ecological Economics*, *116*, 310–321. https://doi.org/10.1016/j.ecolecon.2015.05.001

De Schrijver, E., Folly, C. L., Schneider, R., Royé, D., Franco, O. H., Gasparrini, A., & Vicedo-Cabrera, A. M. (2021). A Comparative Analysis of the Temperature-Mortality Risks Using Different Weather Datasets Across Heterogeneous Regions. *GeoHealth*, *5*(5), e2020GH000363. https://doi.org/10.1029/2020GH000363

Hamilton, L. C., & Stampone, M. D. (2013). Blowin' in the Wind: Short-Term Weather and Belief in Anthropogenic Climate Change. *Weather, Climate, and Society*, *5*(2), 112–119. https://doi.org/10.1175/WCAS-D-12-00048.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hoffmann, R., Muttarak, R., Peisker, J., & Stanig, P. (2022). Climate change experiences raise environmental concerns and promote Green voting. *Nature Climate Change*, *12*(2), 148–155. https://doi.org/10.1038/s41558-021-01263-8

Leiserowitz, A., Maibach, E., Roser-Renouf, C., & Hmielowski, J. (2012). *Global Warming's Six Americas, March 2012 and Nov. 2011*. Yale University and George Mason University, Yale Project on Climate Change Communication. https://climatecommunication.yale.edu/wp-content/uploads/2016/02/2012_07_Global-Warming%E2%80%99s-Six-Americas-in-March-2012-and-November-2011.pdf

Schraufnagel, D. E., Balmes, J. R., Cowl, C. T., De Matteis, S., Jung, S.-H., Mortimer, K., Perez-Padilla, R., Rice, M. B., Riojas-Rodriguez, H., Sood, A., Thurston, G. D., To, T., Vanker, A., & Wuebbles, D. J. (2019). Air Pollution and Noncommunicable Diseases. *Chest*, *155*(2), 417–426. https://doi.org/10.1016/j.chest.2018.10.041

Sikt - Norwegian Agency for Shared Services in Education and Research. (2023a). *EOSC eea-regions, edition 1.0* [dataset]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/EEA-REGIONS

Sikt - Norwegian Agency for Shared Services in Education and Research. (2023b). *EOSC eea-stations, edition 1.0* [dataset]. Sikt - Norwegian Agency for Shared Services in Education and Research   Publication year: (format =) 2022. https://doi.org/10.21338/EEA-STATIONS

Sikt - Norwegian Agency for Shared Services in Education and Research. (2023c). *EOSC era5-grids, edition 1.0* [dataset]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ERA5-GRIDS

Sikt - Norwegian Agency for Shared Services in Education and Research. (2023d). *EOSC era5-regions, edition 1.0* [dataset]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ERA5-REGIONS

Sikt - Norwegian Agency for Shared Services in Education and Research. (2023e). *EOSC Future Science Project "Climate Neutral and Smart Cities"* [dataset]. Sikt - Norwegian

Agency for Shared Services in Education and Research. https://doi.org/10.21338/EOSC-FUTURE-SP9

Sikt - Norwegian Agency for Shared Services in Education and Research. (2023f). *EOSC merged-EOSC-ESS8e02_2, edition 1.0* [dataset]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/MERGED-EOSC-ESS8E02_2

Sikt - Norwegian Agency for Shared Services in Education and Research. (2023g). *EOSC merged-EOSC-ESS9e03_1, edition 1.0* [dataset]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/MERGED-EOSC-ESS9E03_1

Sikt - Norwegian Agency for Shared Services in Education and Research. (2023h). *EOSC merged-EOSC-ESS10, edition 1.0* [dataset]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/MERGED-EOSC-ESS10

Sikt - Norwegian Agency for Shared Services in Education and Research. (2023i). *EOSC merged-EOSC-ESS10SC, edition 1.0* [dataset]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/MERGED-EOSC-ESS10SC

van den Elshou, S. van den E., Bartelds, H., Heich, H., & Léger. (2012). *CiteairII — Common Information to European Air*. DCMR. https://web.archive.org/web/20160222064802/http:/www.airqualitynow.eu/download/CITEAIR-Comparing_Urban_Air_Quality_across_Borders.pdf

Whitmarsh, L. (2008). Are flood victims more concerned about climate change than other people? The role of direct experience in risk perception and behavioural response. *Journal of Risk Research*, *11*(3), 351–374. https://doi.org/10.1080/13669870701552235

Appendix A - Graphical illustration of indices

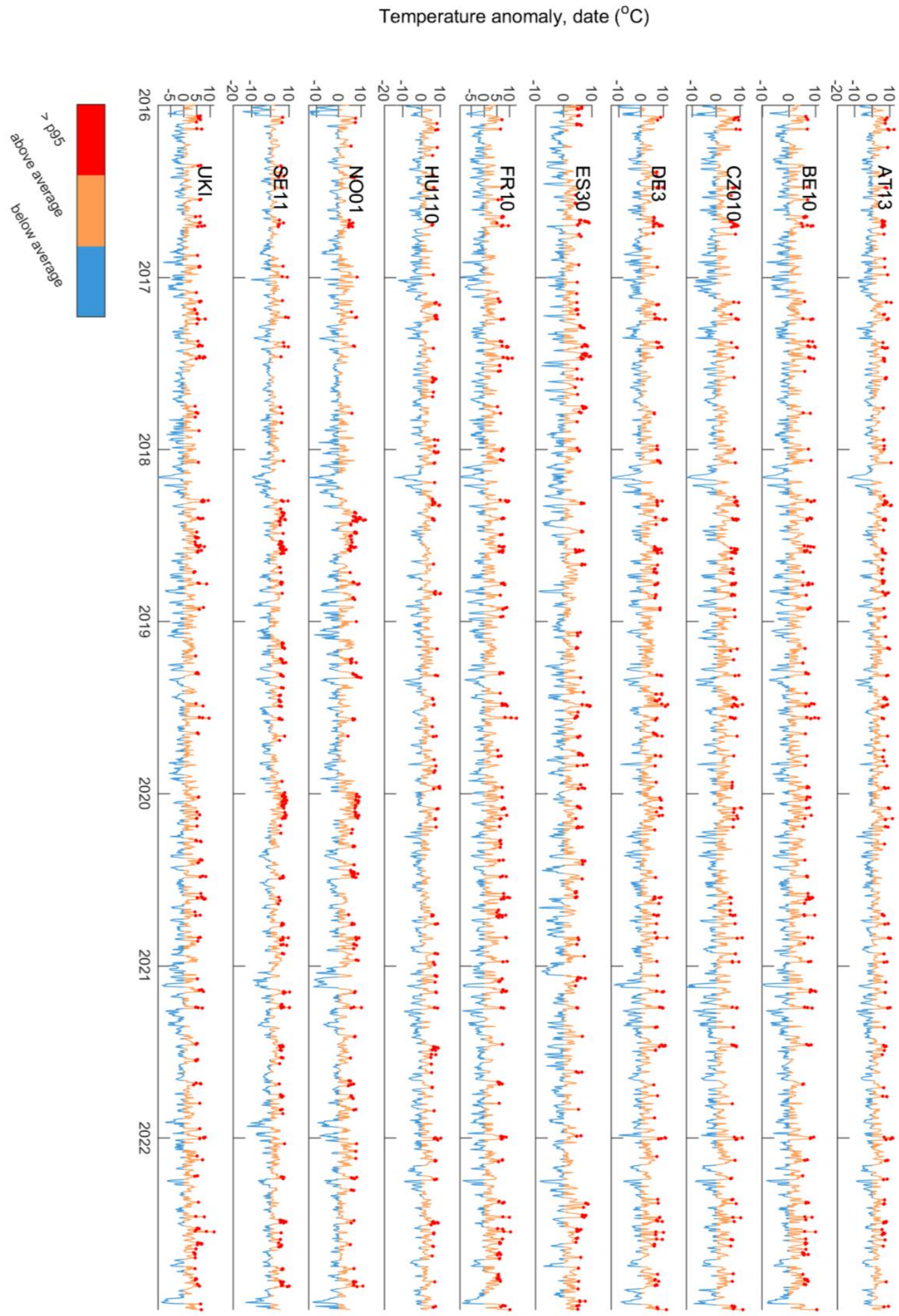

Figure A1: Regional time-series of index **tmpanod (**Temperature anomaly, date), color-coded according to whether the daily values are above or below **tmpdcamb** (temperature in degrees Celcius, multi-year calendar month averages, baseline 1991 - 2020) or above **tmp95pacmb** (Temperature in degrees Celcius, multi-year calendar month 95th percentiles, baseline 1991 - 2020).

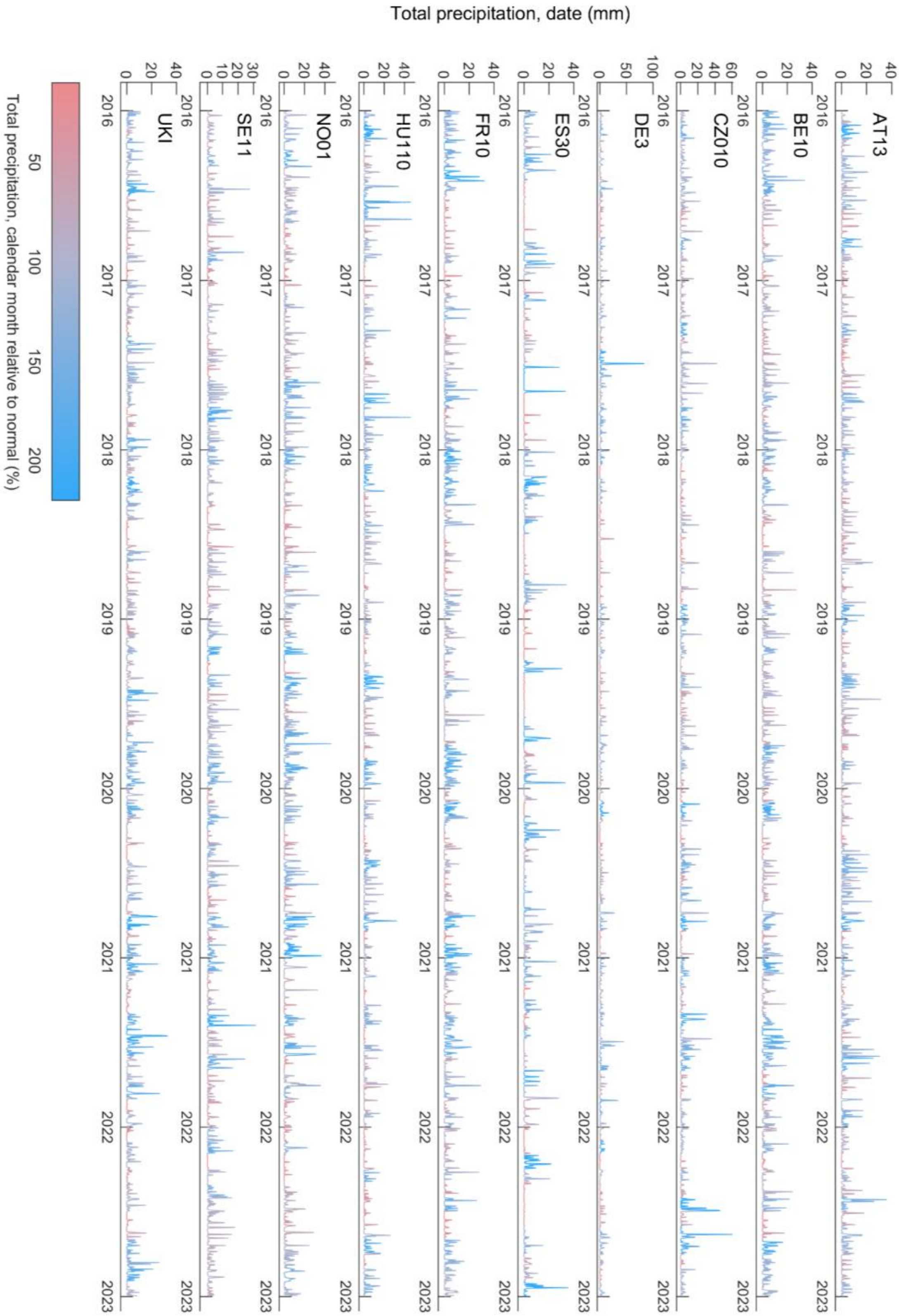

Figure A2: Regional time-series of index **paccta (**Total precipitation average, date), color-coded according to **paccdcm** (Total precipitation, calendar month relative to normal).

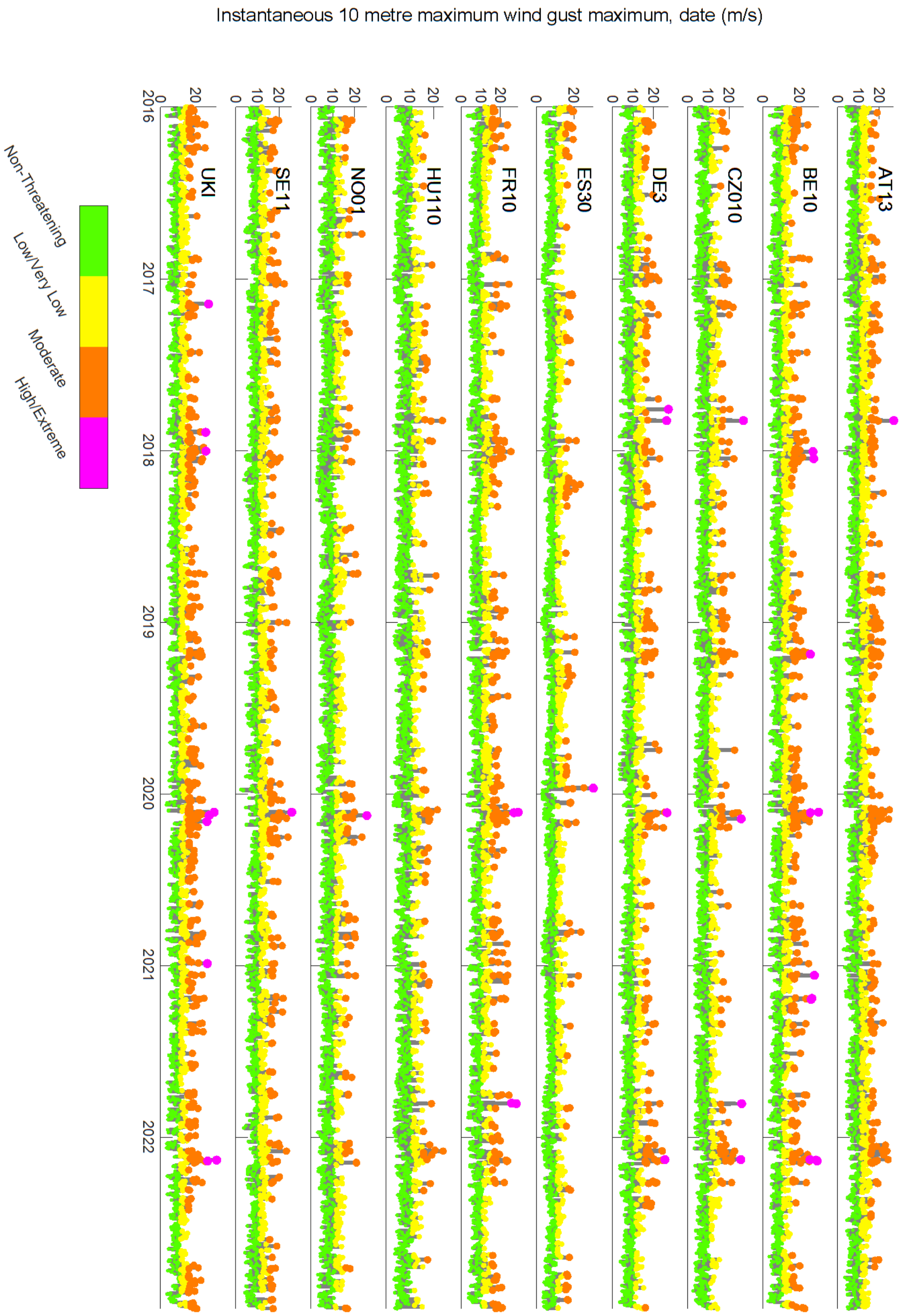

Figure A3. Regional time-series of **iwg10mx** (Instantaneous 10 metre maximum wind gust, date). Wind gust is classified and colored after the National Oceanic and Atmospheric Administration's Wind Threat Description, https://www.weather.gov/mlb/seasonal_wind_threat

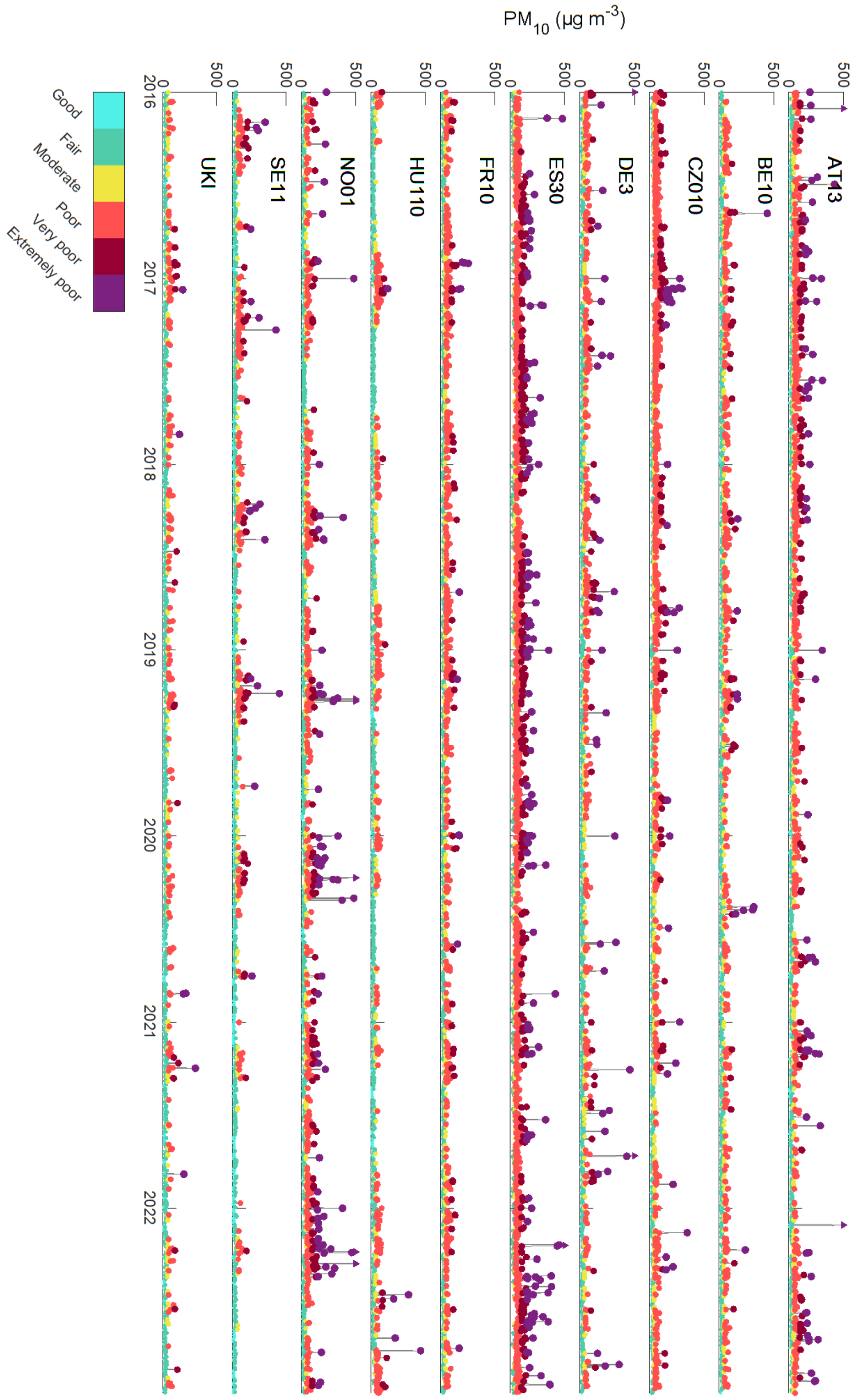

Figure A4: Regional average daily PM_{10} concentrations. Pollutant concentrations are classified and colored according to their Air Quality indices (Table 5 in main text). ▲symbols indicate values exceeding a figure's vertical-axis limit.

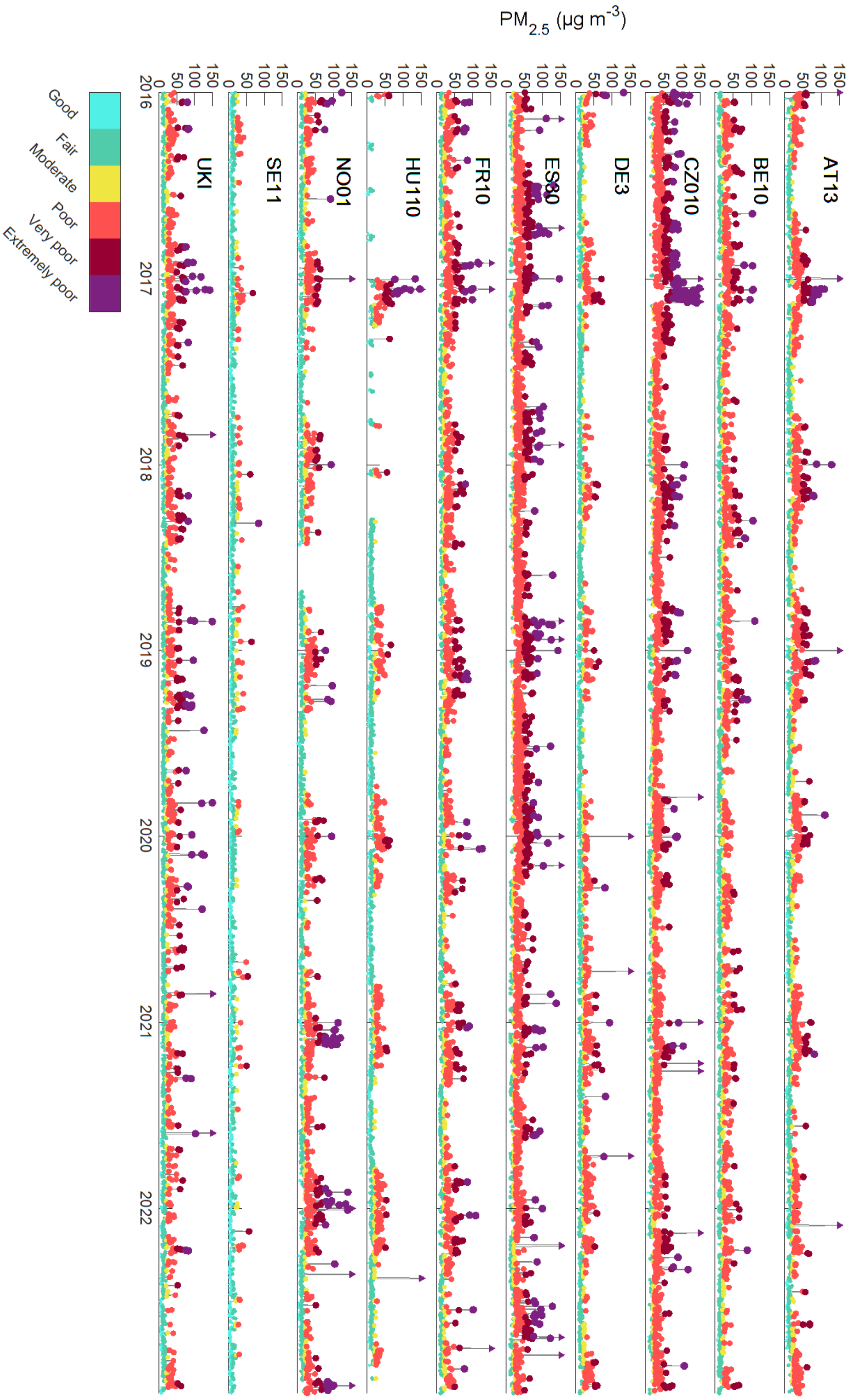

Figure A5: Regional average daily $PM_{2.5}$ concentrations. Pollutant concentrations are classified and colored according to their Air Quality indices (Table 5 in main text). ▲symbols indicate values exceeding a figure's vertical-axis limit.

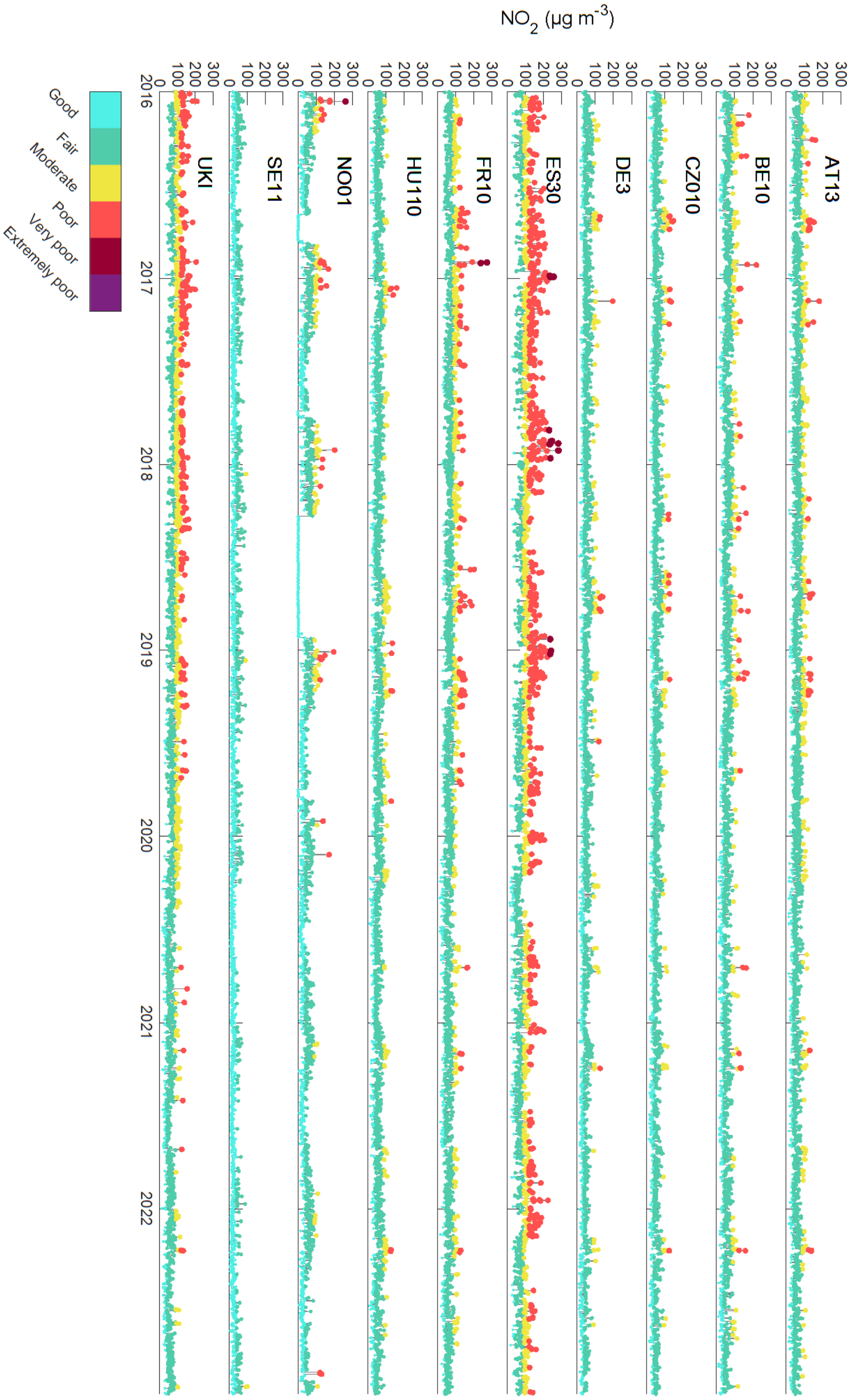

Figure A6: Regional average daily NO_2 concentrations. Pollutant concentrations are classified and colored according to their Air Quality indices (Table 5 in main text). ▲symbols indicate values exceeding a figure's vertical-axis limit.

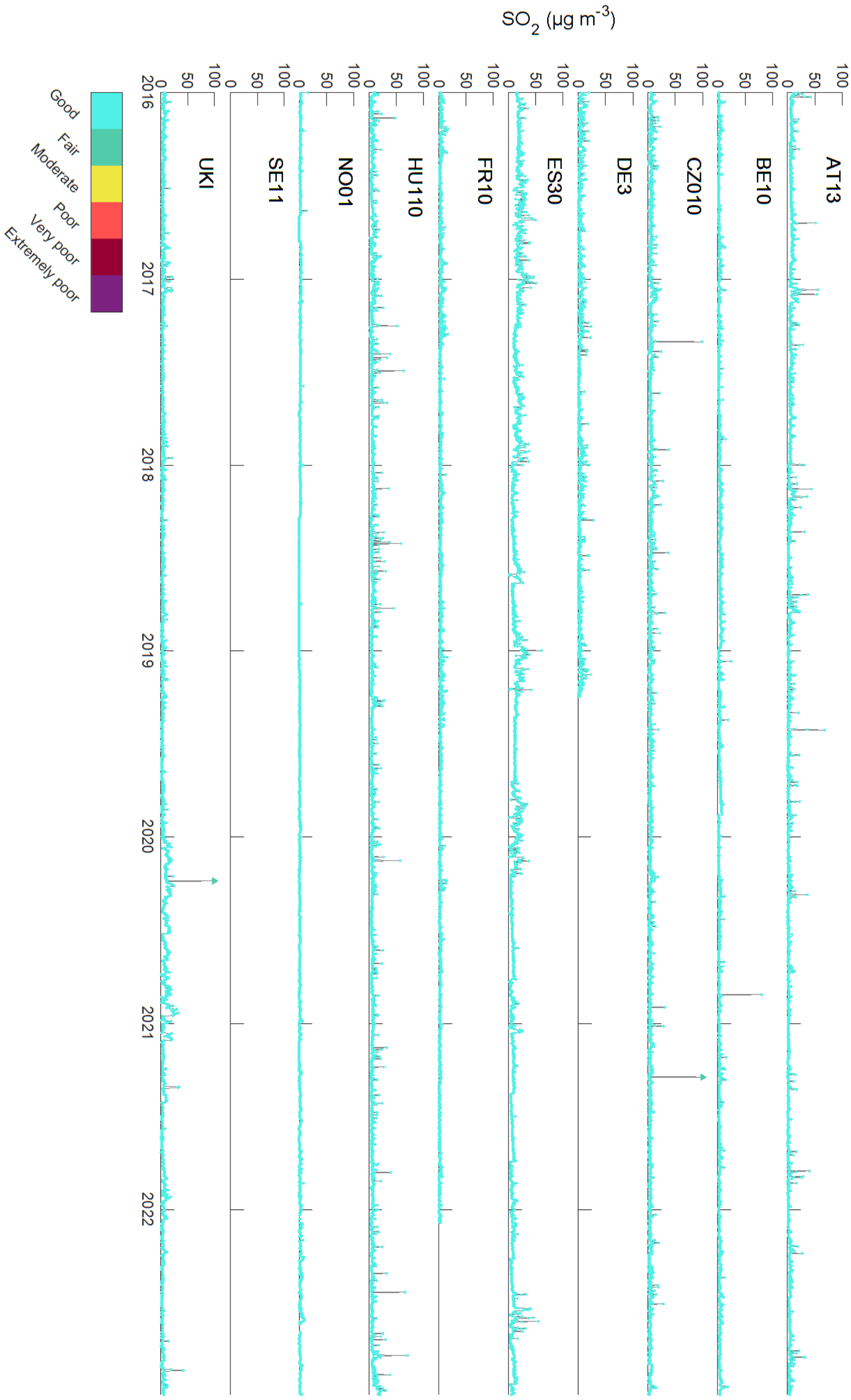

Figure A7: Regional average daily SO_2 concentrations. Pollutant concentrations are classified and colored according to their Air Quality indices (Table 5 in main text). ▲symbols indicate values exceeding a figure's vertical-axis limit.

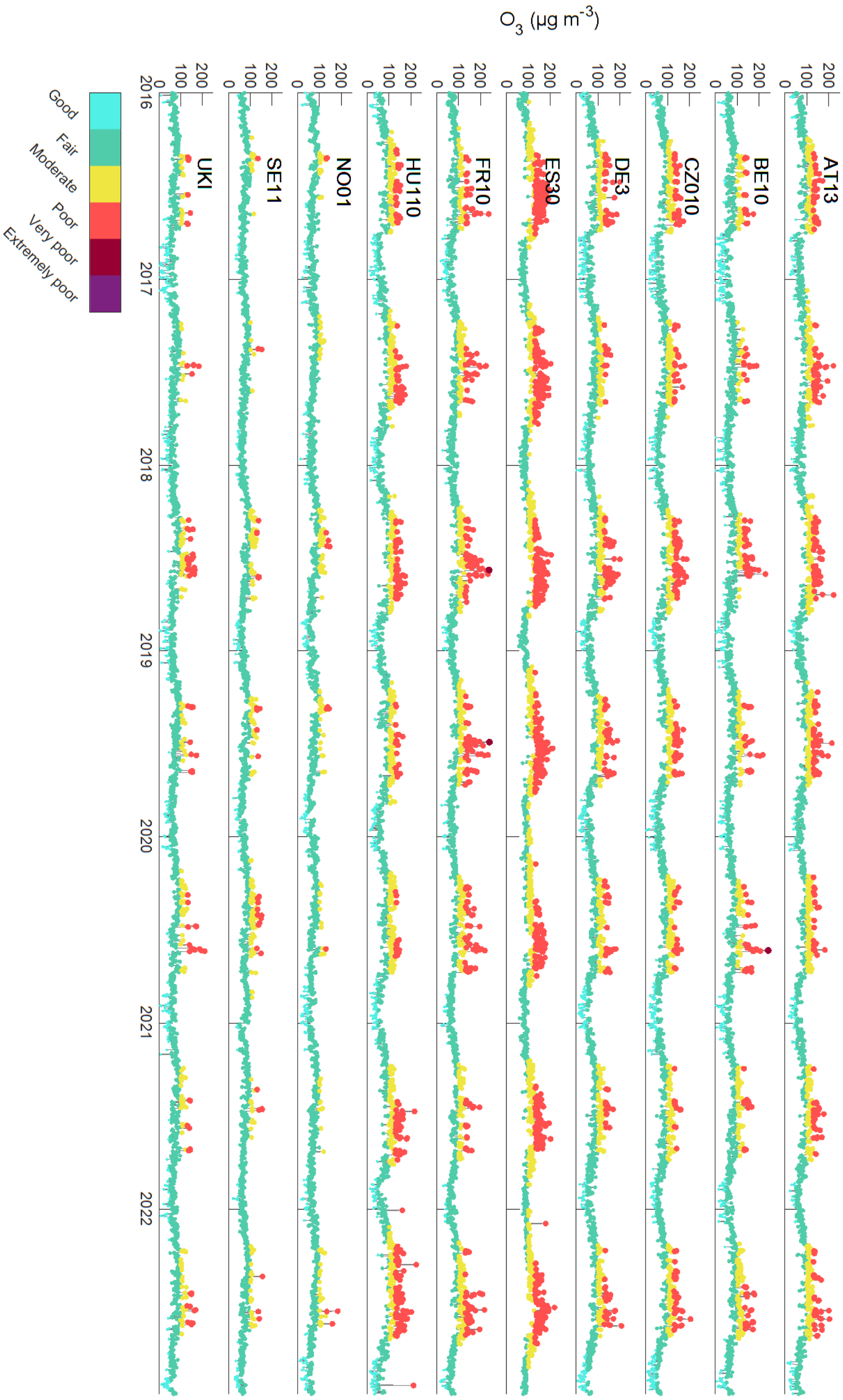

Figure A8: Regional average daily O_3 concentrations. Pollutant concentrations are classified and colored according to their Air Quality indices (Table 5 in main text). ▲symbols indicate values exceeding a figure's vertical-axis limit.