



## Data Article

# An improved and updated 20-years air pollution exposure dataset for Germany 2003-2022 (APEXpose\_DE v6)

Alexandre Caseiro<sup>1,\*</sup>, Erika von Schneidemesser

RIFS Research Institute for Sustainability at GFZ, Germany

## ARTICLE INFO

*Article history:*

Received 14 February 2025

Revised 15 May 2025

Accepted 20 June 2025

Available online 24 June 2025

Dataset link: [APEXpose\\_DE \(Original data\)](#)*Keywords:*

Air quality

Exposure

Health

Germany

## ABSTRACT

The link between exposure to polluted air and the outcome of diseases (e.g., cardio-vascular diseases, COVID-19) has been established. Nevertheless, research on the quantification of the relationship is still relevant today. Quantifying the link between the ambient atmospheric concentrations of pollutants and the outcome of diseases requires knowledge on the levels of the pollutants through time at various scales. In the present work, an improved and updated version of the APEXpose\_DE dataset is described. The dataset provides air pollution metrics at the yearly time resolution and at the spatial resolution of the NUTS-3 level, corresponding to the *Landkreis/Kreisfreie Stadt* in Germany. The dataset evolved from its initial form by expanding the years covered and by refining the gap-filling methodology. The dataset can serve as input to, e.g., observational studies.

© 2025 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

\* Corresponding author.

E-mail address: [Alexandre.Caseiro@external.eumetsat.int](mailto:Alexandre.Caseiro@external.eumetsat.int) (A. Caseiro).<sup>1</sup> Presently at EUMETSAT - European Organisation for the Exploitation of Meteorological Satellites.

## Specifications Table

Subject	Atmospheric Science
Specific subject area	The present dataset provides air pollutant concentration estimates in Germany at the spatial level of the NUTS-3 and at the temporal resolution of one year for the years 2003–2022, relevant for exposure studies.
Type of data	Table
Data collection	Processed
Data source location	Ambient concentrations of atmospheric pollutants measured in Germany in 2003–2022 and reported to the European Environmental Agency were retrieved from the Air Quality Download Service (formerly known as Airbase, under a different architecture). The data was processed and completed with reanalysis data from the Copernicus Atmosphere Monitoring Service, after scaling, in order to estimate the monthly concentrations to which people were subjected to for each county in Germany.
Data source location	The data were collected from the European Environmental Agency through its Air Quality Download Service ( <a href="https://eeadmz1-downloads-webapp.azurewebsites.net/">https://eeadmz1-downloads-webapp.azurewebsites.net/</a> ) and from the Copernicus Atmosphere Monitoring Service through its Atmosphere Data store ( <a href="https://ads.atmosphere.copernicus.eu/datasets/cams-global-reanalysis-eac4?tab=download">https://ads.atmosphere.copernicus.eu/datasets/cams-global-reanalysis-eac4?tab=download</a> ).
Data accessibility	Repository name: Zenodo Data identification number: <a href="https://zenodo.org/records/14873210">10.5281/zenodo.14873210</a> Direct URL to data: <a href="https://zenodo.org/records/14873210">https://zenodo.org/records/14873210</a>
Related research article	<a href="https://doi.org/10.1186/s12940-024-01149-0">10.1186/s12940-024-01149-0</a>

## 1. Value of the Data

- The present data provides long-term air pollution exposure values at a spatial resolution (the NUTS-3 units) which corresponds to health records, or to which health records can easily be aggregated. The temporal resolution of the data is yearly. These three characteristics (temporal and spatial resolutions as well as temporal depth) make the dataset especially useful as input for observational studies, or as a validation dataset for air quality modelling studies.
- Besides its temporal and spatial resolutions as well as its temporal depth, the ASCII format of the provided dataset enables a simple access and workup (e.g., it can be easily rasterized or vectorized).
- The characteristics of the dataset makes it useful for investigating policy; the evolution of exposure to atmospheric pollutants can be evaluated in concert with the evolution of the policy.

## 2. Background

Despite the link between air pollution and adverse health outcomes being well established, high-quality datasets are needed to improve its quantification [1]. The present dataset is an updated and improved version of the dataset described in Caseiro and von Schneidmesser [2]. The first version of the dataset was aimed at supporting observational studies such as the one developed by Koch et al. [3]. The years covered were gradually expanded (through versions 2, 3 and the current one, 6) from 2010–2019 to 2003–2022, now covering 20 years. The downscaling of the CAMS (Copernicus Atmosphere Monitoring Service) reanalysis data was refined through the versions, from a simple linear regression between reanalysis and monitoring-based values (versions 1 and 2), to a multilinear regression further including population, area, and population density of the NUTS-3 units (version 3), to finally including the area type of each NUTS-3 unit as a factor (version 4 onwards). Each NUTS-3 unit can be classified into one of four classes, as defined by the Federal Institute for Research on Building, Urban Affairs and Spatial Development

(Bundesinstitut für Bau- Stadt- und Raumforschung), based on estimations of the population actually present within the unit throughout a typical day: very central, central, peripheral or very peripheral.

### 3. Data Description

The provided dataset comprises the following files:

- APEXpose\_DE\_2003-2022\_changelog.md: a short description of the updates throughout the versions.
- APEXpose\_DE\_2003-2022.README: a short description of the dataset in APEXpose\_DE\_2003-2022\_nogeo.csv (description of the fields, records and metadata).
- APEXpose\_DE\_2003-2022\_nogeo.csv: the air pollution concentration for exposure estimates data itself. Each record (each line in the file) corresponds to a NUTS-3 unit (identified by its name and its code), and a scenario, for a given year. There are 402 NUTS-3 units in Germany and 3 scenarios were developed, the total number of records in the dataset is 1206 per year, or 24120 for the entire study period. Each record includes a numeric value for each metric considered.
- APEXpose\_DE\_2003-2022\_Ratified.csv: describes the level of QA/QC for the original monitoring data.
- APEXpose\_DE\_2003-2022\_StationTypes.csv: describes the station(s) used for each data point, in the case of monitoring data, or indicates the use of reanalysis data.

The variables present in the dataset (file APEXpose\_DE\_2003-2022\_nogeo.csv) are:

- NO<sub>2</sub> annual mean concentration (micrograms per cubic meter)
- number of hours of the year which have an NO<sub>2</sub> concentration over 200 µg per cubic meter
- NO annual mean concentration (micrograms per cubic meter)
- PM10 annual mean concentration (micrograms per cubic meter)
- number of days of the year which have a daily average PM10 concentration over 50 µg per cubic meter
- PM2.5 annual mean concentration (micrograms per cubic meter)
- O<sub>3</sub> annual mean concentration (micrograms per cubic meter)
- number of days of the year which have a daily average O<sub>3</sub> concentration over 120 µg per cubic meter
- annual mean of the daily O<sub>3</sub> maximum concentration (micrograms per cubic meter)
- maximum daily 1 h average O<sub>3</sub> concentration over the entire year (micrograms per cubic meter)
- maximum daily 8 h average O<sub>3</sub> concentration over the entire year (micrograms per cubic meter)

### 4. Experimental Design, Materials and Methods

In all versions of APEXpose\_DE so far, the NUTS-3 units for Germany reflect the status until November 1st, 2016 (402 units), when the two Landkreise Göttingen and Osterode am Harz merged. Two sources were used for the production: (1) measurements from the Air Quality Download Service (formerly Airbase), managed by the European Environmental Agency, and (2) the CAMS global reanalysis EAC4.

#### 4.1. Airbase

The monitoring data, accessed between 2020 and 2023 via the EEA's Air Quality Download Service (formerly, under a different architecture, known as Airbase), was processed in the same

way as described by Caseiro and von Schneidmesser [2] with the exception of expanding to the years 2003–2022. The data sources were as follows:

- 2003–2012: the Airbase-v8 dataset
- 2012–2021: the E1a dataset (ratified)
- 2022: the E2a dataset (unratified)

#### 4.2. CAMS

The yearly averaged CAMS reanalysis data was processed (clipped and vectorized) to be used for gap-filling. In the first version the processed reanalysis data was scaled via a simple linear regression using the NUTS-3 unitwise measurements-based data as dependent variable and the processed reanalysis-based data as independent variable. In the current version multilinear regressions were constructed adding the NUTS-3 population and/or its area and/or its population density and/or its structural settlement type, as defined by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (*Bundesinstitut für Bau- Stadt- und Raumforschung*, [https://www.bbsr.bund.de/BBSR/DE/forschung/raumb Beobachtung/Raumabgrenzungen/deutschland/kreise/Raumtypen2010\\_krs/Raumtypen2010\\_Kreise.html](https://www.bbsr.bund.de/BBSR/DE/forschung/raumb Beobachtung/Raumabgrenzungen/deutschland/kreise/Raumtypen2010_krs/Raumtypen2010_Kreise.html)), as independent variables. The best performing model was selected using the Akaike Information Criterion and used to estimate concentrations for the NUTS-3 units which did not have measurements.

CAMS-based data for the metrics with a time resolution shorter than one day were produced as described in Caseiro and von Schneidmesser [2].

#### 4.3. APExpose\_DE

As a final step, the Airbase- and CAMS-derived data are combined to produce the version 6 of the APExpose\_DE dataset. An example is given in Fig. 1: the NO<sub>2</sub> annual mean for 2021 (average scenario).

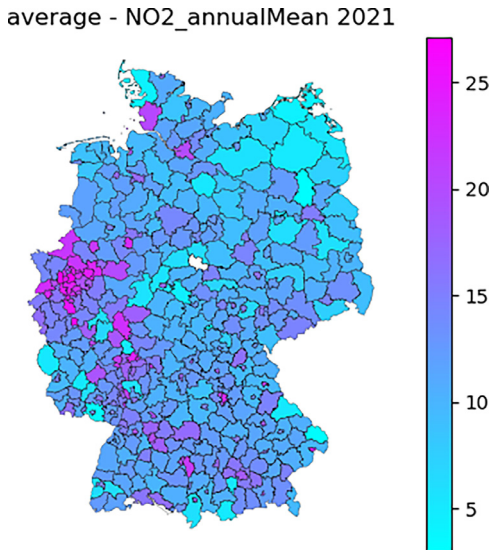


Fig. 1. annual average NO<sub>2</sub> concentration ( $\mu\text{g.m}^{-3}$ ) for 2021.

## Limitations

The main drawback of the dataset is the limited number of NUTS-3 units with background stations (at best about half of the 402 NUTS-3 units), and the consequent need for gap filling using reanalysis data. Despite the refined scaling strategy used, this approach introduces a methodological discontinuity, whose implications must be investigated in, e.g., observational studies.

## Ethics Statement

The authors have read and follow the ethical requirements for publication in *Data in Brief*, confirming that the work presented in this article does not involve human subjects, animal experiments, or any data collected from social media platforms.

## Data Availability

[APEXpose\\_DE \(Original data\)](#) (Zenodo)

## CRediT Author Statement

**Alexandre Caseiro:** Conceptualization, Resources, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization; **Erika von Schneidmesser:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Acknowledgments

We wish to acknowledge the RIFS Research Institute for Sustainability at GFZ for funding the research of Erika von Schneidmesser and Alexandre Caseiro with financial support provided by the Federal Ministry of Education and Research of Germany (BMBF) and the Ministry for Science, Research and Culture of the State of Brandenburg (MWFK).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] A. Pozzer, S.C. Anenberg, S. Dey, A. Haines, J. Lelieveld, S. Chowdhury, Mortality attributable to ambient air pollution: a review of global estimates, *GeoHealth* 7 (2023) e2022GH000711, doi:[10.1029/2022GH000711](https://doi.org/10.1029/2022GH000711).
- [2] A. Caseiro, E. von Schneidmesser, APEXpose\_DE, an air quality exposure dataset for Germany 2010–2019, *Sci. Data* 8 (2021) 287, doi:[10.1038/s41597-021-01068-6](https://doi.org/10.1038/s41597-021-01068-6).
- [3] S. Koch, C. Hoffmann, A. Caseiro, M. Ledebur, M. Menk, E. von Schneidmesser, Air quality in Germany as a contributing factor to morbidity from COVID-19, *Environ. Res.* 214 (Part 2) (2022) 113896, doi:[10.1016/j.envres.2022.113896](https://doi.org/10.1016/j.envres.2022.113896).