Data in brief 28 (2020) 105089

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Spatio-temporal data on the air pollutant nitrogen dioxide derived from Sentinel satellite for France

Hichem Omrani [a], Bilel Omrani [b,c], Benoit Parmentier [d], Marco Helbich [e,*]

[a] Urban Development and Mobility, Luxembourg Institute of Socio-Economic Research, Luxembourg
[b] École Centrale de Lyon, France
[c] Polytechnique Montréal, Canada
[d] University of Maryland, National Socio-Environmental Synthesis Center, USA
[e] Department of Human Geography and Spatial Planning, Utrecht University, Utrecht, Netherlands

ARTICLE INFO

Article history:
Received 11 December 2019
Received in revised form 25 December 2019
Accepted 27 December 2019
Available online 7 January 2020

Keywords:
Air pollution
Remote sensing
Monitoring

ABSTRACT

Monitoring of air pollution is an important task in public health. Availability of data is often hindered by the paucity of the ground monitoring station network. We present here a new spatio-temporal dataset collected and processed from the Sentinel 5P remote sensing platform. As an example application, we applied the full workflow to process measurements of nitrogen dioxide (NO_2) collected over the territory of mainland France from May 2018 to June 2019. The data stack generated is daily measurements at a 4 × 7 km spatial resolution. The supplementary *Python* code package used to collect and process the data is made publicly available. The dataset provided in this article is of value for policy-makers and health assessment.

* Corresponding author.
E-mail address: M.Helbich@uu.nl (M. Helbich).

https://doi.org/10.1016/j.dib.2019.105089

Specifications Table

Subject	Environmental science
Specific subject area	Pollution
Type of data	Table Image Dataset Python code
How data were acquired	The data was collected via the Sentinel Hub API.
Data format	Processed raw data Analyzed Filtered
Parameters for data collection	The daily air pollution data was collected from May 2018 to June 2019.
Description of data collection	All the data was obtained from the Sentinel 5P satellite using the application programming interface. The collected data are processed by a *Python* code.
Data source location	France
Data accessibility	Data is supplied on Mendeley (Public repository) Repository name: https://doi.org/10.17632/zxwc456xy9.5

Value of the Data

- The collected air pollution data offer a (cost-)effective and scalable source for advancing the monitoring of NO_2 at a large scale.
- Policy makers can generate yearly, monthly, or even near real-time datasets on air pollution maps to support government efforts in reducing air pollution.
- The automated workflow to process Sentinel 5P data is transferable to any other study area on the globe.
- Satellite data can be merged with ground-based measurement or survey data.

1. Data description

Remote sensing data for air quality monitoring is important for health research [1]. The advantage of remotely sensed air pollution data includes, but is not limited to, large coverage at a useful spatial and temporal resolution. Sentinel 5P is a rather new remote sensing data source but requires downloading and computationally intensive processing that is often a barrier to public use. For illustrative purposes, we focus on generation of a NO_2 spatio-temporal measurements across mainland France. The shared workflow repository, however, contains other air pollutants including ozone (O_3), sulfur dioxide (SO_2), carbon monoxide (CO) (for details, see the GitHub page [2]). The spatial resolution of the measurements (3.5×7 km^2 for all trace gases, except for CO and Methane (CH_4) that is 7×7 km^2) allows observations and mapping of air pollution at a finer scale (e.g. at the scale of an administrative area) (see Table 1 and Fig. 5).

The data consist of a netCDF file containing Sentinel 5P measurements between May 2018 and June 2019, with multiple attributes (e.g., latitude/longitude, WGS84 projection, and date of measurement), allowing both spatial and temporal observations of air pollutants. We cleaned the data using a quality flag parameter (noted by 'qa_value' varying between 0 (no data) and 1 (full quality data). We used 'qa_value' above 0.5 provided by the Copernicus Sentinel 5 Precursor Tropospheric Monitoring Instrument (S5p/TROPOMI [3]) to filter cloud cover and so to ensure high quality data. The dense measurements performed every 24 hours allow accurate annual averaging at fine-grained spatial resolution as shown in Fig. 1.

The data were also grouped by date allowing temporal assessment of pollutants density and quantitative observations such as monthly pollutants distribution, both spatially and temporally (Fig. 2). Data allow also analysis for weekday variations (Fig. 3).

Table 1
Average tropospheric vertical column of NO_2 by administrative area.

Administrative area	Average tropospheric vertical column of NO_2 ($\times 10^{19}$ molec/m^2)
Corse	1.782
Île-de-France	3.923
Hauts-de-France	3.345
Nouvelle Aquitaine	1.670
Normandie	2.445
Pays de la Loire	1.985
Centre-Val de Loire	2.208
Grand Est	2.913
Provences-Alpes-Côtes d'Azur	2.170
Bretagne	1.922
Bourgogne-France-Comté	2.174
Occitanie	1.659
Auvergne-Rhône-Alpes	1.951

Note: Row data is publicly accessible on Mendeley repository [4].

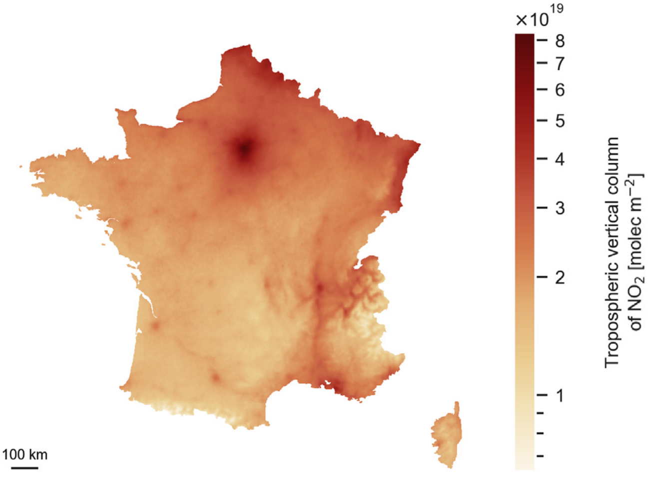

Fig. 1. Annual average of NO_2 concentrations in France between May 2018 and June 2019.

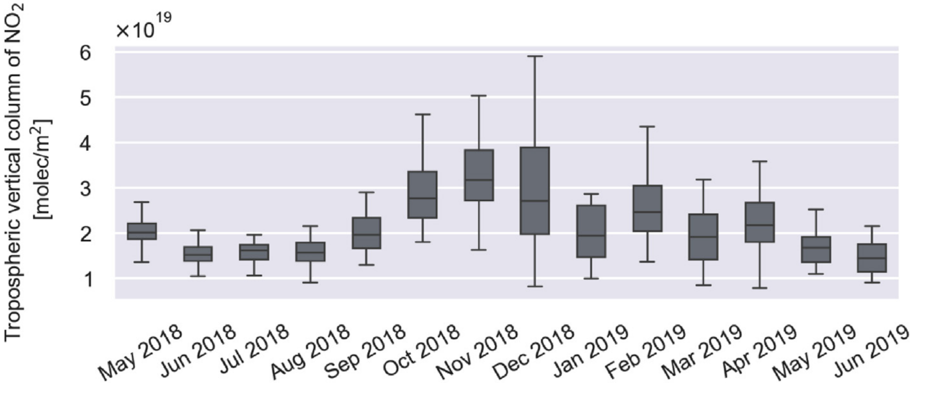

Fig. 2. Distribution of NO_2 measurements collected over France between May 2018 and June 2019.

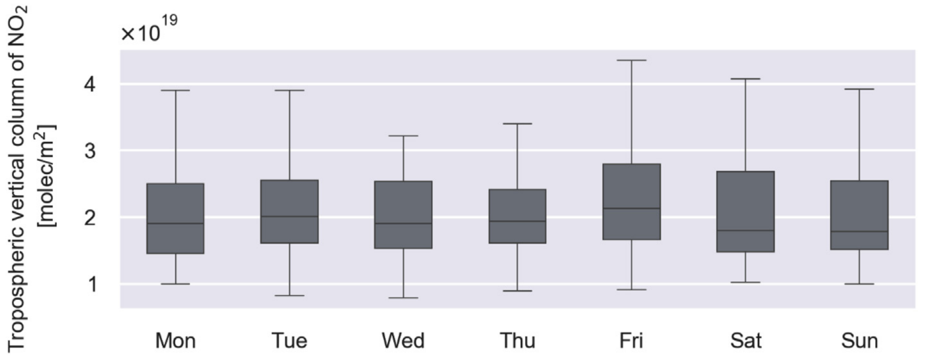

Fig. 3. Distribution NO_2 measurements collected over France by weekday and weekend.

Fig. 4 illustrates that the seasons fall and winter (October to February) face higher NO_2 pollutants in the Northern and Eastern part of France. This spatial contrast reduces to a point where only city urban areas show strong pollution (e.g., June 2019 with Paris).

2. Experimental design, materials, and methods

1) Satellite measurements

Launched in October 2017 by the European Space Agency (ESA), Copernicus Sentinel 5P [5] monitors the density of several atmospheric gases, aerosols, and cloud distributions affecting air quality and climate. The measurements are made by the state of the art instrument called TROPOspheric Monitoring Instrument (TROPOMI). The TROPOMI is a multispectral imaging spectrometer that detects solar radiation reflected or scattered back to space from Earth's atmosphere and surface. As the spectral fingerprint of each target atmospheric trace gas is known, its concentration can be calculated through the identification of the unique fingerprints of these constituents in different part of the electromagnetic spectrum. Sentinel 5P is able to achieve global coverage every 24 hours, giving access to dense measurements over the entire globe. TROPOMI has more spectral bands than its predecessors: ultraviolet and visible (270–500 nm), near-infrared (675-77 nm), and shortwave infrared (2305–2385 nm). This allows TROPOMI to measure a wider range of atmospheric trace gases such as nitrogen dioxide (NO_2), ozone (O_3), sulphur dioxide (SO_2), methane (CH_4), and carbon monoxide (CO). In addition, it observes clouds and aerosols-related parameters, which can be fed into the retrieval algorithms of trace gases [3]. The list of standard S5P/TROPOMI L2 products is given in Table 2.

Overall, the dataset provides:

- geolocated total columns of ozone, sulfur dioxide, nitrogen dioxide, carbon monoxide, formaldehyde and methane-geolocated cloud and absorbing aerosol index
- other products are under development and will made available at a later date. They include geolocated tropospheric columns of ozone, geolocated vertical profiles of ozone, aerosol layer height, ultraviolet index, etc.

All of the mission's measurements of atmospheric gases and aerosols are ‘column data’, which means they cover the full depth of the atmosphere. For some gases, advanced techniques and algorithms like ozone profile retrieval, the convective cloud differential, and the cloud slicing methods allow to have access to ‘tropospheric column densities’, ‘stratospheric column densities’ and ‘vertical density profiles’. When available, these variables are included in the dataset.

TROPOMI has a very high spatial resolution (3.5×7 km^2 for all trace gases, except for CO and CH_4 that is 7×7 km^2). Further, TROPOMI has an improved signal-to-noise ratio (2–5%) for measurements

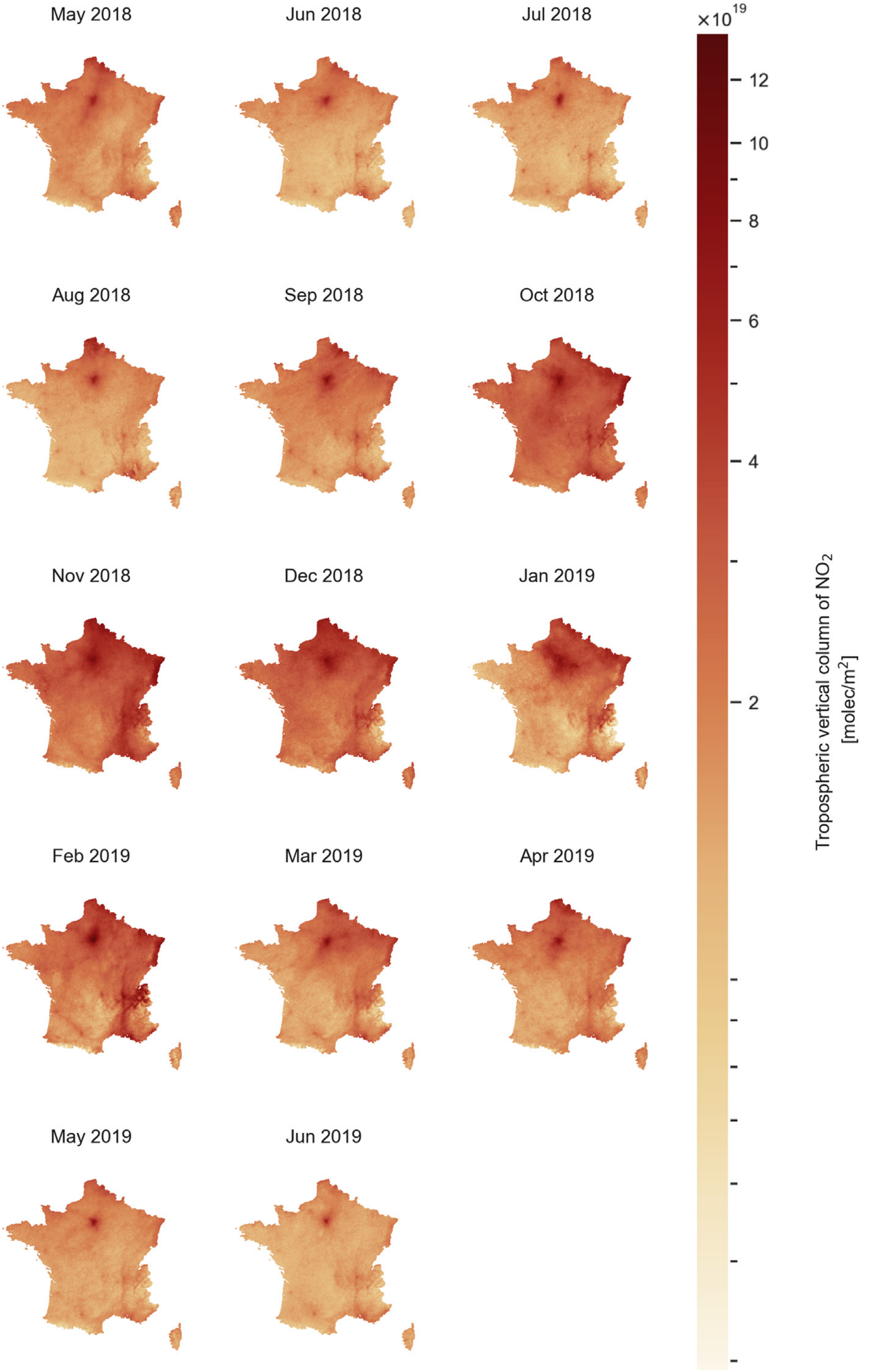

Fig. 4. Spatial distribution of NO_2 measurements across France by month.

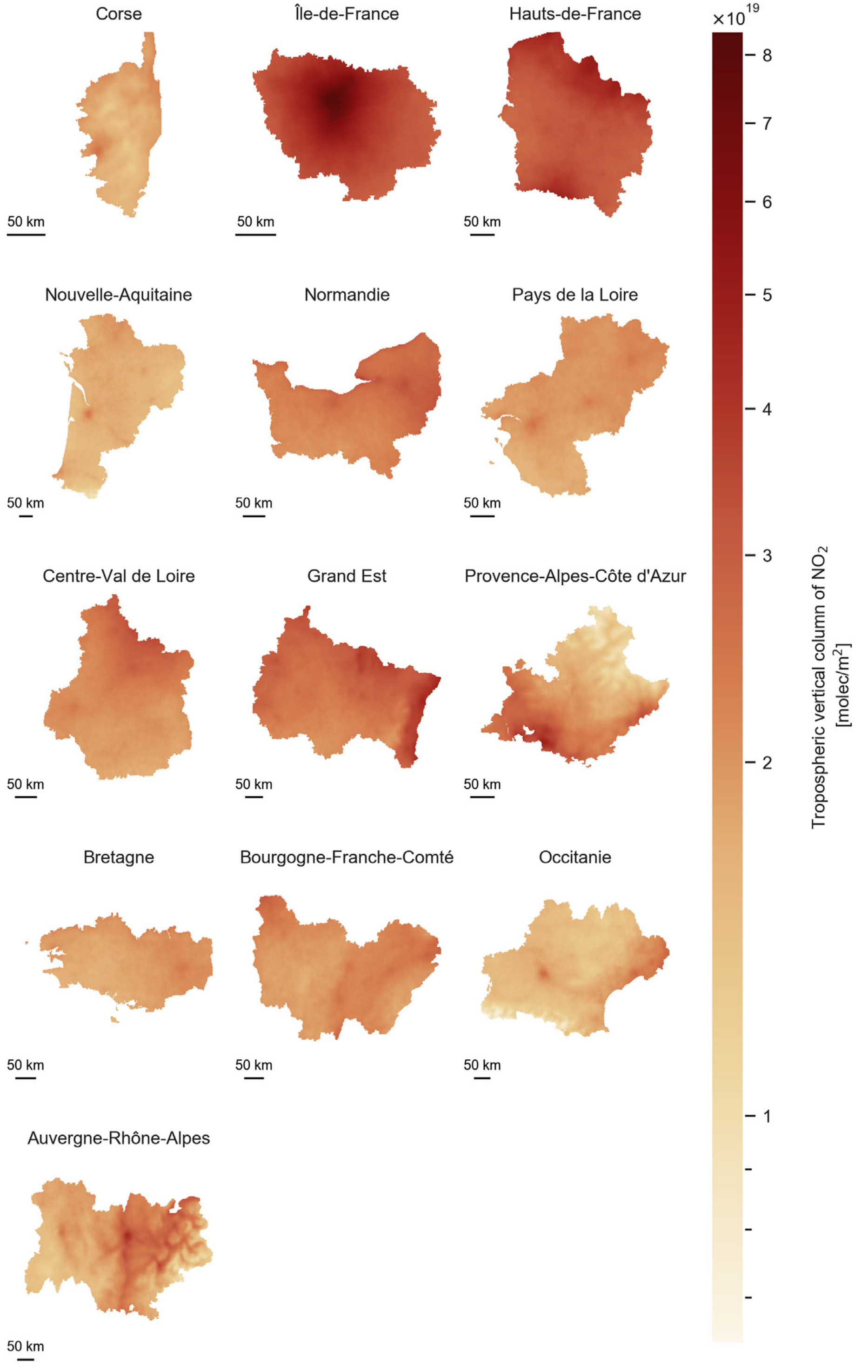

Fig. 5. Annual average of NO2 measured between May 2018 and June 2019 by administrative area.

Table 2
List of S5P/TROPOMI level 2 data products.

Product	Main Parameter	(Planned) released
UV aerosol index	Aerosol index	Released
Aerosol layer height	Mid-level pressure	Mid-2019
Carbon monoxide (CO)	Total column	Released
Cloud	Fraction, albedo, top pressure	Released
Formaldehyde (HCHO)	Total column	Released
Methane (CH_4)	Total column	Released
Nitrogen dioxide (NO_2)	Total, tropospheric, stratospheric column	Released
Ozone profiles	Total and tropospheric profiles	Late-2019
Sulfur dioxide (SO_2)	Total column	Released
Ozone (O_3)	Total column	Released
Tropospheric ozone (O_3)	Tropospheric column	In development
Ultraviolet (UV)	Surface irradiance erythemal dose	In development

under low albedo conditions. Data gaps are documented by Copernicus and can be consulted on the mission page.

2) Processing workflow

For further analysis of TROPOMI data, we produced an aggregated product on a regular grid with spatial resolution of 3.5×7 km (0.01×0.01 arc degrees). Every orbit that TROPOMI measures has a different spatial distribution of grid cells, depending on the viewing zenith angle at the moment of the observation. For this reason, we resampled each product for the area of interest on this single grid and binned the dataset by latitude/longitude WGS84 projection. The quality of the individual observations depends on many factors, including cloud cover, surface albedo, presence of snow-ice, saturation, geometry etc. In Sentinel 5P, a layer summarizing the different factors affecting the quality of the measurements is provided. This aggregate measure called 'quality assurance value' ('*qa_value*') can be used to screen poor quality pixel. This '*qa_value'* is a continuous variable, ranging from 0 (no date) to 1 (all is well). To filter out errors and problematic retrievals we excluded measurements with a 'qa_value' < 0.5 following the Copernicus specifications [3].

Acknowledgments

The authors are grateful to the Luxembourg Institute of Socio-Economic Research (LISER) for funding the internship of Bilel Omrani, École Centrale de Lyon-France and Polytechnique Montréal-Canada, for data collection and processing with Python. The authors would also like to acknowledge the European Spatial Agency for providing the API for the Sentinel 5P Hub. The content is solely the responsibility of the authors and does not necessarily represent the official views of the LISER.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.105089.

Conflict of Interests

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] V.V. Putrenko, N.M. Pashynska, The use of remote sensing data for modeling air quality in the cities, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017, p. 57, 4.
[2] Sentinel-5P. https://github.com/bilelomrani1/s5p-tools. (Accessed July 2019).

[3] Copernicus. https://www.copernicus.eu/en. (Accessed 10 June 2019).
[4] Row data is publically available on the following repository. https://drive.google.com/drive/folders/1t5vbQq1g0LtJa37Sc6NYoq45bkLP2EWp?usp=sharing.
[5] J.P. Veefkind, et al., TROPOMI on the ESA Sentinel-5 Precursor: a GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote Sens. Environ. 120 (SI) (2012) 70–83, https://doi.org/10.1016/j.rse.2011.09.027.