Data in Brief 49 (2023) 109398

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

# Air quality low-cost sensors and monitoring stations $NO_2$ raw dataset in Rouen (France)

Emma Thulliez[a,*], Bruno Portier[a], Michel Bobbia[b,*], Jean-Michel Poggi[c,d]

[a] INSA Rouen Normandie, Normandie Univ, LMI UR 3226, F-76000 Rouen, France
[b] ATMO Normandie, 3 place de la pomme d'or, 76000 Rouen, France
[c] LMO, Université Paris-Saclay, Fac. Sciences Orsay, Bat. 307, 91400 Orsay, France
[d] Dept SD, IUT Paris-Rives de Seine, Université Paris Cité, Paris, France

ARTICLE INFO

*Article history:*
Received 5 June 2023
Revised 28 June 2023
Accepted 7 July 2023
Available online 11 July 2023

Dataset link: Air quality low-cost sensors and monitoring stations NO2 raw dataset in Rouen (France) (Original data)

*Keywords:*
Micro-sensors
Nitrogen dioxide
AirSensEUR
Air Pollution
Sensor calibration

ABSTRACT

This article presents a dataset comprising measurements made by co-located devices, with the aim of calibrating sensors for an upcoming in-situ use. The dataset includes hourly averaged data from 9 low-cost sensors and 2 traffic monitoring stations (thereafter named QDP and SUD3) in Rouen spanning from October 20, 2021 to March 25, 2022. In addition, the dataset is enriched by covariates measured by the sensors: temperature, relative humidity, atmospheric pressure, plus Ox and CO measures.
The experiment was conducted as part of TIGA[1]'s call for project, and designed to have a better understanding of sensors' drawbacks, particularly when they are moved or shut down. Knowledge about the effect of air pollution on health has gained significant attention from both the scientific community and citizens, making air quality a growing issue for urban area. As a result, the city of Rouen in Normandy, France, has prioritized air quality monitoring as a key initiative. Concurrently, several means to measure air pollutants have been made more accessible, such as the use of low-cost sensors. Those sensors offer affordability, but are known to

* Corresponding authors.
*E-mail addresses:* emma.thulliez@insa-rouen.fr (E. Thulliez), michel.bobbia@atmonormandie.fr (M. Bobbia).

[1] TIGA: Territoire d'Innovation Grande Ambition (see https://www.banquedesterritoires.fr/territoires-dinnovation in French). It consists of a program to improve innovation in France, which allow resources for 24 cities in this aim. The project carried by the city of Rouen is to enhance fluidity in transportation, and reduce its environmental impact. Eventually, Rouen would propose to travel by autonomous bus for an everyday ride.

https://doi.org/10.1016/j.dib.2023.109398

be less accurate than monitoring stations. Thus, they need to be cautiously studied so as to be used properly.



## Specifications Table

| | |
|---|---|
| Subject | Environmental Science (air pollution) |
| Specific subject area | Nitrogen dioxide ($NO_2$) low-cost sensors calibration using monitoring stations |
| Type of data | Table (CSV files) |
| How the data were acquired | Instruments connected to the ATMO Normandie[2] network:<br>• Low-cost sensors: 9 Alphasense Chemical Sensors in AirSensEUR host (including NO2-B43F, NO-B4, CO-A4, OX-A431 and deported temperature, humidity and pressure sensors in each node, as detailed in [1]). As described in [2,3], AirSensEUR is an Open Platform project, targeted to precise air quality measurements, developed by the Joint Research Center in Ispra, in collaboration with Liberaintentio S.r.l.<br>• Reference measures: 2 HORIBA APNA 370 (chemiluminescence) ambient NOx monitors, installed at traffic sites (named SUD3 and QDP). The QDP one also provides CO measurements. |
| Data format | Hourly averaged raw data |
| Description of data collection | Data were collected by ATMO Normandie using 9 low-cost sensors and 2 monitoring stations located in Rouen, France, close to major roads. Data are acquired every 15 minutes for monitoring stations and every minute for low-cost sensors, and transmitted via the cellular network. However, the dataset only includes measurements averaged over 1 hour as they are the most often used. Data were collected starting from October 20, 2021 to March 25, 2022, thus making a total of 29,605 observations. This period is particularly interesting since it corresponds the highest $NO_2$ concentrations in a year in Rouen.<br>Reference measurements are given in µg/m$^3$ while data from the sensors are non-calibrated, that is to say they represent electrical potential difference in volts. Temperature is given in Celsius degrees, relative humidity in percent and pressure in hPa (or equivalently in mbar).<br>At the beginning of the measurements campaign, each sensor is installed co-located with one traffic monitoring station (4 at the SUD3 site and 5 at the QDP one). Most of them were permuted to the other one so as to study the stability of the calibration models when the device is turned off, moved and installed in a different location. |
| Data source location | Institution: ATMO Normandie<br>City/Town/Region: Rouen<br>Country: France<br>Latitude and longitude for collected samples/data:<br>- At QDP: longitude: 1.098583 latitude: 49.43675<br>- At SUD3: longitude: 1.0670139 latitude: 49.4341639<br>*As GPS coordinates:*<br>- *At QDP: longitude: 1°05’54.90”E latitude: 49°26’12.30”N*<br>- *At SUD3: longitude: 1°04’01.25”E latitude: 49°26’02.99”N* |
| Data accessibility | Repository name: Mendeley Data<br>Data identification number: 10.17632/82dnstrd93.1<br>Direct URL to data: http://dx.doi.org/10.17632/82dnstrd93.1 |

[2] See https://www.atmonormandie.fr/

**Value of the Data**

- This dataset enables air quality monitoring organization to calibrate a set of $NO_2$ low-cost sensors and understand the variables involved in the bias of measurements.
- As such, this dataset can help researchers in developing new models to calibrate low-cost sensors. This is as helpful for the scientific community as for citizens or companies interested by these technologies for their own use or for sensor-makers who usually lack in-situ measurements.
- By the design of the experiment, this dataset enables researchers to study the influence of moving or shutting down a sensor for a short or long period of time over its calibration; but also study its possible temporal drift.
- This dataset can be used to quantify the variability between low-cost sensors from the same models.
- This dataset can be used to study the assets of having multiple collocated $NO_2$ sensors.

## 1. Objective

As part of TIGA's call for projects, the Rouen Normandy metropolis presented a research plan to improve mobility and transportation around the city center. The project includes the idea of creating a more fluid transport network, but also of reducing the environmental impact of the current and upcoming network. As such, monitoring $NO_2$ emissions on major roads became necessary – since they are mainly caused by traffic, as it was confirmed again by the EEA[3] in 2022 (see [4]).

Fifty low-cost sensors, which are more affordable with respect to reference monitoring devices (up to a hundred times less expensive) are to be installed in Rouen and its surroundings by ATMO Normandie, the Air Quality Monitoring Agency of Normandy. However, prior to being deployed in real-world settings, those sensors must be calibrated thanks to more reliable monitoring stations (see [5] for the need of in-situ calibration). Our dataset includes the measurements made by a subset of sensors and their associated monitoring devices over a period of five months, corresponding to high $NO_2$ concentrations. This dataset is supposed to help calibrating low-cost sensors in Rouen, so that ATMO Normandie can make use of more $NO_2$ measurement sites and improve air quality cartography.

## 2. Data Description

The dataset is divided into 9 CSV files (one per sensor) named 'ASExx.csv' where 'xx' denotes the sensor's id. ('xx' goes from 4 to 13, excluding 9). Each file is organized as follows, having for columns:

- date: the timestamp in UTC, in the format yyyy-mm-dd HH:MM where dd denotes the day, mm the month, yyyy the year, HH the hour and MM the minutes.
- ASE_NO2: the hourly averaged concentration of $NO_2$ measured by the sensor in µV.
- ASE_NO: the hourly averaged concentration of NO in µV.
- ASE_CO: the hourly averaged concentration of CO in µV.
- ASE_Ox: the hourly averaged concentration of Ox in µV.
- ASE_T: the hourly averaged temperature inside the node in°C.
- ASE_HR: the hourly averaged relative humidity inside the node in %.
- ASE_PA: the hourly averaged atmospheric pressure in hPa.
- SUD3_NO2: the hourly averaged concentration of $NO_2$ measured by the reference monitor SUD3 in $\mu g/m^3$.

[3] European Environment Agency (see https://www.eea.europa.eu/en)

**Table 1**
Localization of the sensors depending on the time period and number of observations gathered.

| Low-cost sensor | Localization | Period | Number of observations |
|---|---|---|---|
| ASE4 | SUD3 | 2021-10-21 to 2021-12-01 | 983 |
| | QDP | 2021-12-13 to 2022-03-25 | 2445 |
| ASE5 | SUD3 | 2021-10-20 to 2022-01-24 | 2302 |
| | QDP | 2022-01-24 to 2022-03-25 | 1436 |
| ASE6 | SUD3 | 2021-10-21 to 2022-03-25 | 3720 |
| ASE7 | QDP | 2021-10-26 to 2022-01-24 | 2161 |
| | SUD3 | 2022-01-24 to 2022-03-25 | 1438 |
| ASE8 | Shut down[4], QDP | 2021-10-25 to 2022-03-04 | 3117 |
| | | 2022-03-04 to 2022-03-25 | 239 |
| ASE10 | SUD3 | 2021-10-20 to 2021-12-01 | 1008 |
| | QDP | 2021-12-01 to 2022-01-24 | 1291 |
| | SUD3 | 2022-03-14 to 2022-03-25 | 1437 |
| ASE11 | QDP | 2021-12-16 to 2022-03-04 | 1871 |
| ASE12 | QDP | 2021-10-25 to 2021-12-01 | 883 |
| | SUD3 | 2021-12-13 to 2022-03-25 | 2447 |
| ASE13 | QDP | 2021-11-05 to 2021-12-01 | 626 |
| | SUD3 | 2021-12-02 to 2022-01-24 | 1269 |
| | QDP | 2022-01-24 to 2022-03-25 | 932 |

- SUD3_NO: the hourly averaged concentration of NO measured by the reference monitor SUD3 in µg/$m^3$.
- QDP_NO: the hourly averaged concentration of NO measured by the reference monitor QDP in µg/$m^3$.
- QDP_NO2: the hourly averaged concentration of $NO_2$ measured by the reference monitor QDP in µg/$m^3$.
- QDP_CO: the hourly averaged concentration of CO measured by the reference monitor QDP in µg/$m^3$.
- location: Either QDP or SUD3, the place where the node was located at the time of the measurements.

Each row of a file represents an observation of the pollutants, averaged over the last hour (excepts for the date and location). Each missing measurements is represented by a 'NA' (Not Available).

Table 1 lists the main information to know for each node, that is where they were located, at which time, and how many observations it represents. The sum of the number of rows of each file leads up to a total of 29,605 observations.

As it can be noticed from Table 1, sensors ASE6 and ASE11 were never moved from their first localization, one at each monitoring station. Sensor ASE8 wasn't moved neither, but was shortly shut down on purpose to give this situation coverage. This lasted a few seconds. Sensors ASE4, ASE5, ASE7 and ASE12 were placed at a monitoring station during a first period prior to being moved to the other one. Eventually, sensor ASE10 was installed at the QDP monitor, then moved to the SUD3 one, and brought back at QDP. Another two-ways moving scenario was applied to sensor ASE13, which was at first installed at SUD3.

It should be noted that there are at least 3 days of missing observations from sensor ASE6. This didn't occur on purpose, and it is likely that the device was turned off during this period, thus giving an example of an extended shutdown.

[4] The sensor ASE8 has been shut down at 10am, leading to NA values on the 3117$^{th}$ observation (2022-03-04 10:00).

Fig. 1 shows the scatter plots of monitoring station measurements against sensor measurements for two given sensors. This gives an example of the difference in behavior from one sensor to another. Fig. 2 illustrates how covariates from our dataset is an added-value; in this case for fitting a multiple linear regression model. The model used was estimated with 469 observations.

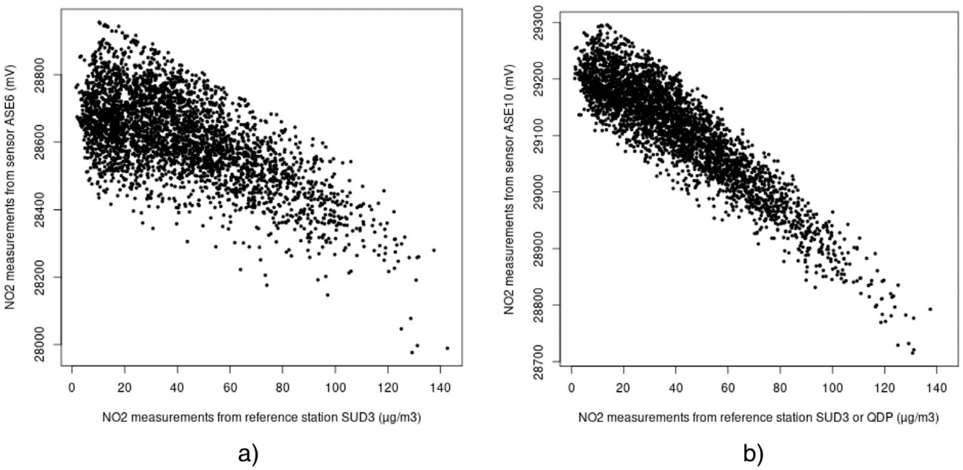


**Fig. 1.** Scatter plots of measurements from reference stations against co-located low-cost sensors. a) Monitoring station SUD3 v. Sensor ASE6 (2021-10-21 to 2022-03-25) b) Monitoring station SUD3 v. Sensor ASE10 (2021-10-20 to 2021-12-01 and 2022-03-14 to 2022-03-25) or QDP v. ASE10 (2021-12-01 to 2022-01-24).

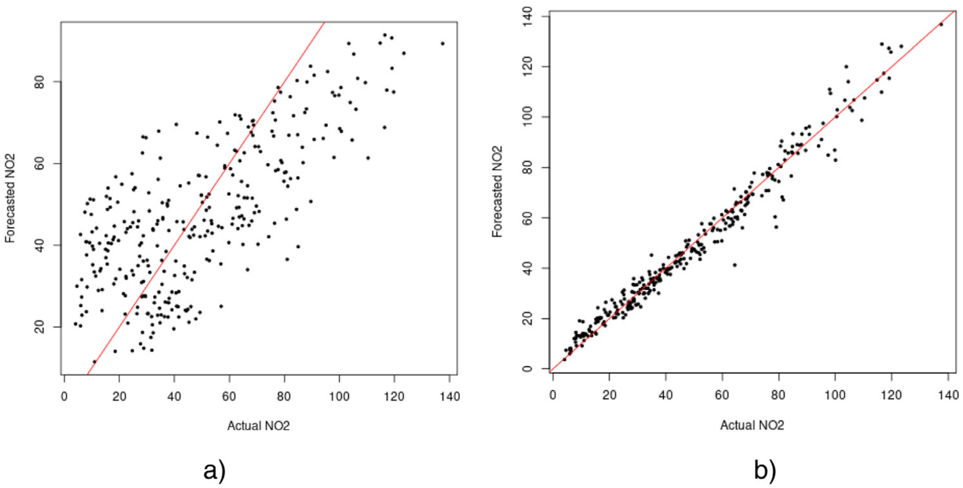


**Fig. 2.** Scatter plots of actual against forecasted $NO_2$ measurements by a) a simple linear regression and b) a multiple linear regression model making use of other measurements given by the sensor ASE6.

## 3. Experimental Design, Materials and Methods

Sensors have been installed and monitored by ATMO Normandie, which is the organization owning the high-quality devices. They have been located at the QDP and SUD3 stations, which are used to evaluate NOx concentration in Rouen (see Fig. 3).

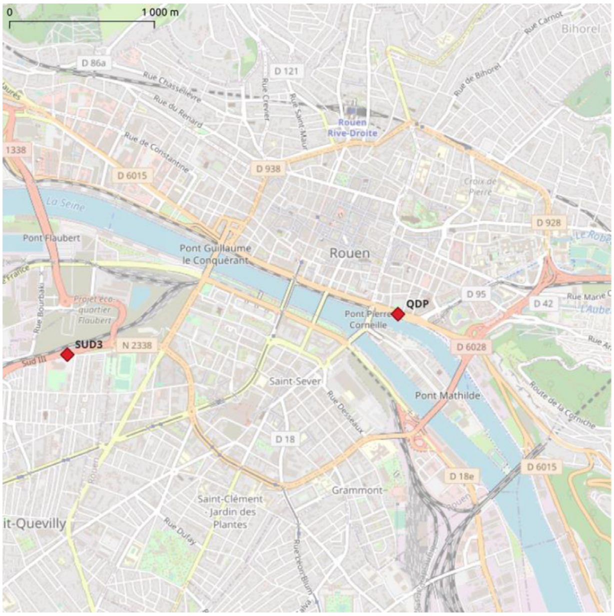


**Fig. 3.** Location of stations QDP and SUD3 over a close view of Rouen. Main roads are shown in red and orange.

The dataset includes measurements made by the first 9 Alphasense Chemical Sensors in AirSensEUR host that were installed in the city. Those measurements were taken as part of the in-situ calibration phase, since: 1) it is recommended to calibrate low-cost sensors in-situ because of the change in environment in contrast with the lab in which they were made (see for example [5]) 2) the measurements made by the AirSensEUR are acquired in µV and not µg/m$^3$ which is the SI base unit.

ATMO Normandie manages multiple monitoring stations in Rouen. The SUD3 and QDP ones (model HORIBA APNA 370, making use of chemiluminescence technology) were chosen as reference because they are considered as traffic station while the other ones measure background concentrations. Since this experimentation focuses on NOx, it is of the most importance to calibrate the sensors near roads, where NOx concentrations are higher.

The QDP site is located at 1°05E and 49°26N on the quay boarding the *Seine* River. This quay is one of the most taken by cars in Rouen. On the other hand, the SUD3 station is placed near a national road leading to the highway, at 1°04E and 49°26N.

Reference monitors measure NO and $NO_2$ in µg/m$^3$. The QDP monitor also provides CO measurements in µg/m$^3$. On the other hand, each sensor node is equipped to measure NO, $NO_2$, CO, Ox (in µV), temperature inside the node (in Celsius degree), relative humidity in %, and atmospheric pressure in hPa.

Each sensor sends its measurements at a frequency of 100 Hz to the manufacturer, as indicated in [3]. Those measurements are then averaged on every minute and send to a cloud platform. On the other hand, reference measurements are available every 15 minutes. For those two networks, measurements were averaged hourly, since this is the norm for broadcasting such data.

This experimentation lasted from October 10, 2021 to March 25, 2022. Over this period, 29,605 observations were collected from low-cost sensors.

The choice of sensor position and duration at each site (shown in Table 1) was motivated by multiple objectives. Firstly, our main goal was to calibrate each sensor by means of a model which should provide a $NO_2$ measurement in µg/m$^3$. Then, each model was to be tested against another monitoring station, so as to evaluate robustness of the calibration after a change in environment. In this aim, some sensors were also moved a second time to make new measurements on the site they were calibrated. Secondly, two sensors were shutdown: one for a few seconds, and the other for a few days. The objective was to test how a sensor could behave after being stopped or fixed, and how it impacts the efficiency of calibration models. However, two sensors were never moved from their original location, making it possible to study the existence of a temporal drift. Eventually, co-locating several sensors had for goal to evaluate inter-sensor variability under different environments (the two sites) and different time periods.

## Ethics Statements

The study does not involve experiments on humans or animals.

## Declaration of Competing 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.

## Data Availability

Air quality low-cost sensors and monitoring stations NO2 raw dataset in Rouen (France) (Original data) (Mendeley Data).

## CRediT Author Statement

**Emma Thulliez:** Writing – original draft; **Bruno Portier:** Conceptualization, Supervision; **Michel Bobbia:** Resources, Data curation; **Jean-Michel Poggi:** Writing – review & editing.

## Acknowledgments

Funding: ATMO Normandie's project is co-financed by the "Rouen mobilité intelligente pour tous" program led by the Métropole Rouen Normandie (winner of the French government's 3rd investment programme (PIA3) "Territoires d'innovation" (TIGA)) and by the European Regional Development Fund (ERDF) of the Normandy Region.

Moreover, we would like to thank Véronique Delmas, head of ATMO Normandie, for providing the data.

## References

[1] M. Gerboles, L. Spinelle, M. SignoriniInstitute for Environment and Sustainability, AirSensEUR: an open data/software/hardware multi-sensor platform for air quality monitoring. Part A: sensor shield, Publications Office, 2015, doi:10.2788/30927.

[2] M. Gerboles, L. Spinelle, M. Signorini, A. KotsevInstitute for Environment and Sustainability, AirSensEUR: an open data/software /hardware multi-sensor platform for air quality monitoring - Part B: Host, influx datapush and assembling of AirSensEUR, Publication Office, 2016, doi:10.2790/214743.

[3] A. Kotsev, S. Schade, M. Craglia, M. Gerboles, L. Spinelle, M. Signorini, Next generation air quality platform: openness and interoperability for the internet of things, Sensors 16 (2016) 403, doi:10.3390/s16030403.
[4] European Environment AgencyAir Quality in Europe 2022, Online Publication, 2022 Report no. 05/2022, doi:10.2800/488115.
[5] C. Borrego A.M. Costa, J. Ginja, M. Amorim, M. Coutinho, K. Karatzas, Th.Sioumisc N. Katsifarakis, K. Konstantinidis, S. De Vito, E. Esposito, P. Smith, N. Andre, P. Gerard, L.A. Francis, N. Castell, P. Schneider, M. Viana, M.C. Minguillon, W. Reimringer, R.P. Otjes, O. von Sicard, R. Pohle, B. Elen, D. Suriano, V. Pfister, M. Prato, S. Dipinto, M. Penza, Assessment of air quality microsensors versus reference methods: TheEuNetAir joint exercise, Atmos. Environ. 147 (2016) 246–263, doi:10.1016/j.atmosenv.2016.09.050.