# ON THE ORGANIZATION AND VALIDATION OF A PILOT TEST OF A MOBILE CROWDSOURCED AIR QUALITY MONITORING SYSTEM

D. Garcia, F. Vázquez-Gallego and M. E. Parés[*]

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA)
Av. Carl Friedrich Gauss 7, Building B4, 08860 Castelldefels, Spain
{danielly.garcia, francisco.vazquez, eulalia.pares}@cttc.es

**KEY WORDS:** Urban modelling, crowdsourcing, air quality maps, smart cities.

**ABSTRACT:**

The development of new tools that allow continuous monitoring of air quality is essential for the study of actions, in order to improve the levels of pollutants in the air that are harmful to the health of citizens. Cardiovascular and respiratory diseases have been identified as risk factors for death in patients with COVID-19; at the same time, exposure to air pollution is associated with these diseases. In this article, we present the pilot tests of the Crowdsourced Air Quality Monitoring (C-AQM) system, which allows the generation of reliable air pollution maps, using data provided by low-cost sensor nodes. The results verify that the system is correct after performing a data calibration; an improvement in $NO_2$ pollution has been observed on weekends, as well as a situation of less air pollution by $NO_2$ between the first and second pandemic waves in Spain.

## 1. INTRODUCTION

A recent study of the Institute of Global Health of Barcelona (ISGlobal) reveals that, only in Spain, $NO_2$ is responsible for more than 9,150 premature deaths (Khomenko et al., 2021). Almost at the same time, last November 2020, the European Commission announced that 30% of the EU funds for 2021-2027 will be spent to fight climate change, the highest share ever of the largest European budget ever. Thus, at last, climate change and air pollution are considered a threat to European society. Technicians are working hard to improve cities' air quality, and traffic restrictions in the main European cities are a clear example. Another extended action is to define healthy paths connecting the main city spotlights.

A key tool to properly carry out these objectives is to know about the actual situation by means of reliable pollution measurements. The Copernicus European system provides maps with high coverage and spatial sampling 7x7 km$^2$ around Europe (ESA, 2017), however, this resolution does not allow us to identify the levels of air pollution in different streets of a neighborhood, Figure 1. The European Commission provides information on the current air quality situation, based on measurements carried out at more than 2000 air quality measurement stations across Europe. In Spain, the control and surveillance of air quality are carried out through the Networks of the Autonomous Communities and Networks of local entities. Thus, national and regional authorities are using maps generated from the interpolation of measurements acquired at reference stations located in relevant points of the territory among others (Jiménez et al., 2008).

Low-cost sensors for measuring air pollution are attracting increasing attention, offering air pollution monitoring at a cost one hundred times less than conventional methods and in theory, making air pollution monitoring possible in many more locations (Gerboles et al., 2017). We will briefly summarize some of the most recent technological advances in air quality
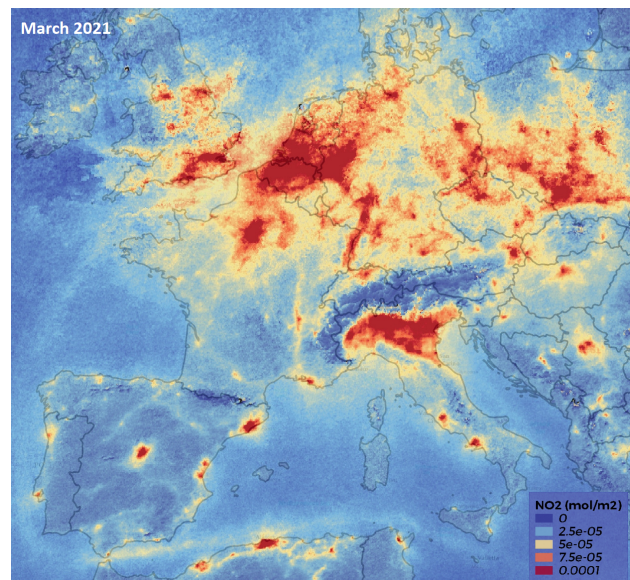


Figure 1. $NO_2$ levels in Europe in March 2021. ©DEFIS_EU

control and management methods, at a lower cost. There are several categories of inexpensive sensors currently available:

- Electrochemical sensors: they are based on a chemical reaction between gases in the air and the electrode within a sensor.

- Photo ionization detector: ionizes volatile organic compounds and measures the result of electrical current.

- Optical particle counters: detect pollution particles, measuring the light scattered by the particles.

- Optical sensors: detect gases such as carbon monoxide and carbon dioxide by measuring the absorption of infrared light.

---

* Corresponding author

Measurements with inexpensive sensors are often of lower and more questionable data quality than the results from official stations. Sensor signals not only depend on the polluting air of interest, but also on a combination of various effects, such as other interfering compounds, temperature, humidity, pressure and signal drift (signal instability). Therefore, the quality of sensor results depends on technology and implementation (application, site, conditions, configuration). Nevertheless, in certain situations, the measurement uncertainty of these devices can approach the level of "official" measurement methods (Gerboles et al., 2017).

The latest developments in air monitoring sensors (Castell et al., 2017) and their price reduction opens the door to the deployment of dense wireless networks of low-cost sensors that provide useful and reliable measurements based on the use of the geodetic principle of redundancy. Even using low-grade sensors, the use of high amount of them could provide better results (Parés et al., 2020). In our first work (Parés and Vázquez-Gallego, 2018), we proposed the Crowdsourced Air Quality Monitoring (C-AQM) system, which relies on the measurements obtained by reference stations, and a cluster of low-cost and low-energy sensor nodes to generate high-resolution air quality maps. In this paper, we will present the first pilot tests of the C-AQM system. Firstly, a brief description of the system is presented. Then, the main organizational aspects are detailed, followed by a summary of the results. Finally, our interpretation of the results, together with project conclusions are presented.

## 2. OVERVIEW OF THE C-AQM SYSTEM

The design and operation of the C-AQM system is detailed in (Parés et al., 2020) Hereafter a brief description of the system and its operation is described for the sake of completeness. The architecture of the C-AQM system is based on three subsystems as shown in Figure 2: 1) acquisition subsystem; 2) processing subsystem; and 3) storage, analysis and visualization subsystem.
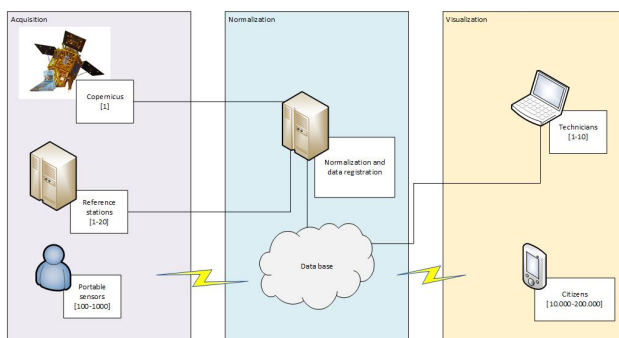


Figure 2. Architecture of the Crowd-sourced Air Quality Monitoring System.

The operation of the C-AQM system is briefly described as follows. Each user or data provider of C-AQM will install on his/her (non-motorized) vehicle a low-cost sensor node, called AirCrowd (Air Quality Crowd-sourced sensing device), Figure 3, developed in our previous work (Parés and Vázquez-Gallego, 2018). The AirCrowd device collects measurements from its sensors and transmits the data to the Cloud through a Narrow Band-IoT (NB-IoT) cellular network. All data collected from multiple AirCrowd devices is stored in an InfluxDB

database, which also includes information provided by reference stations. Then, an office operator can download the data from the database, correct the geo-referenciation, calibrate the system, generate the results and finally, present them through maps and a web viewer of geographic data. The functionalities of each subsystem are described in detail in (Parés and Vázquez-Gallego, 2018). Hereafter, a brief description is done.
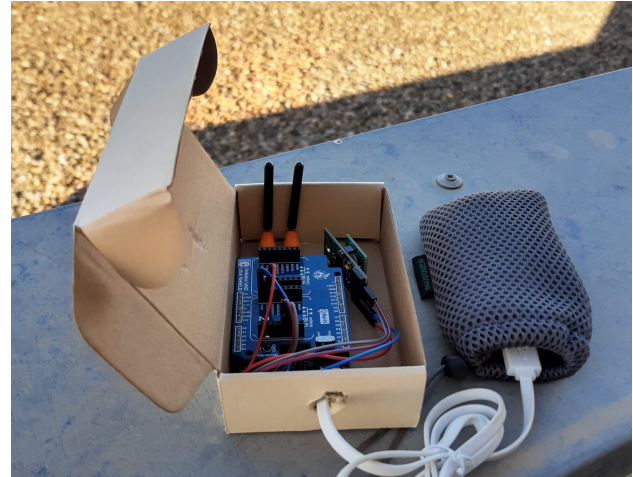


Figure 3. AirCrowd devices of C-AQM.

### 2.1 Acquisition and Processing Subsystem

The acquisition subsystem is the one in charge of collecting data, both air quality measurements and positioning data. The acquisition subsystem is based on two different data sources: 1) official air quality monitoring reference stations, and 2) a network of low-cost sensor nodes.

AirCrowd sensor nodes implement three basic functionalities: (1) acquisition of measurements from gas and particle sensors; (2) acquisition of data from GPS receiver; and (3) wireless transmission of air quality measurements and positioning data to the Cloud. The AirCrowd sensor node was designed as a low-cost, long lifetime, battery-powered, light-weight and small form-factor portable device.

The processing subsystem is the one in charge of processing the raw data in order to provide reliable calibrated and geo-located pollutant measurements. In order to ensure a proper geo-referencing of all collected data, the C-AQM system uses information from the positioning sensors included in the Air-Crowd sensor nodes (i.e., GNSS single frequency receiver) and Open Street Maps.

In the execution of the processing subsystem, three main files are generated. The first file maintains the original coordinate of the input data and groups the data from the sensors and the official reference station into a single file. The data of the second generated file has the coordinates corrected in relation to the reference coordinates. To generate the last file, an algorithm is used for distributed calibration of $NO_2$ measurements, with data from official air quality measurement stations.

### 2.2 Storage, Analysis and Visualization Subsystems

The storage, analysis and visualization subsystem is the one in charge of storing and providing the information to users in a

friendly and understandable way. Regarding the storage functionality, we have selected InfluxDB[1] because it is reported as the current most popular time series database and allows the visualization of the data in real time, using the Grafana platform.

In order to enable the user to easily understand differences, similarities and to allow correlations between the information represented, two ways of displaying the data were developed. On the one hand, maps were generated using Geographic Information System (GIS) software. On the other hand, a web viewer of a Web Map Service (WMS) implemented from the data was created.

The web viewer was prepared using the Spatial Data Infrastructure (SDI) standards and concepts, fulfilling a series of interoperability conditions. The publication of the data through the viewer was divided into two stages. The first was the preparation of the data using the QGIS program and the implementation of the WMS service using GeoServer. The second stage was the creation of an html document for the creation of the web environment and the call to the data layers of the WMS services. In this phase, the OpenLayers library (implemented with JavaScript) was used.

## 3. PILOT OF THE C-AQM SYSTEM

The objectives defined for the pilot test of the C-AQM system were: carry out the validation of the C-AQM system with real data, analyze the air quality in the study area and observe the interest of the citizens in the project. To achieve this, initially, data capture campaigns were carried out with the help of volunteers and later, the processing and analysis of the results, which are described in the following sections.

### 3.1 Pilot design

The pilot tests took place at the city of Sabadell (Spain) and eleven measurement campaigns were performed. All of them were conducted in the north of the city, in an area of around 2 $km_2$. In order to ensure maximum measured area, six different walking routes were designed (Figure 4). Each volunteer involved in the pilot has been recommended to follow one of the walking routes (Figure 5). To enable subsequent data processing, all routes started and finished very close to one of the air quality reference stations of the city.

The official reference station for the measurement of air quality in the city of Sabadell is located in such a way that the levels of pollutants measured are mainly determined by the emissions from vehicles on a nearby street or highway. Data on the main environmental pollutants that are harmful to people's health are measured, processed by computer and made available to users hourly. The levels of the pollutants measured by the reference station were used in the calibration stage of the C-AQM measurements.

The routes were designed with the following criteria: have of around 45 minutes, the sensors had to cross during the route and cover the most representative streets of the neighborhood (i.e., streets with a lot of traffic, no traffic, wide avenues, narrow streets, squares, etc.). The campaigns were always conducted at rush hours (morning or afternoon) both on working days and on

---

[1] https://www.influxdata.com/

weekends, to make more obvious the influence of road traffic on pollution.

To start the data capture, the sensors were connected to the batteries and their correct operation was verified on the Grafana platform, which allows real-time visualization of the data capture. The sensors and maps of the routes were then distributed to the volunteers. Each one was walking along the defined route and with the sensor in a bag. It is important to highlight that the cardboard box allows the entry of air into the $NO_2$ gas sensor, as well as the material of the bags given to the volunteers (Figure 5).

At the end of each data collection campaign, the impressions of the volunteers about the environmental situation during the campaigns were also collected to be later compared with the actual situation. Half of the tests were done pre-pandemic, and the other were done between the first pandemic wave and the second wave in Spain, in September 2020 (Table 1). The data acquired were processed and represented on printed maps and in a web data viewer. Then, the results were presented to the volunteers for feedback.
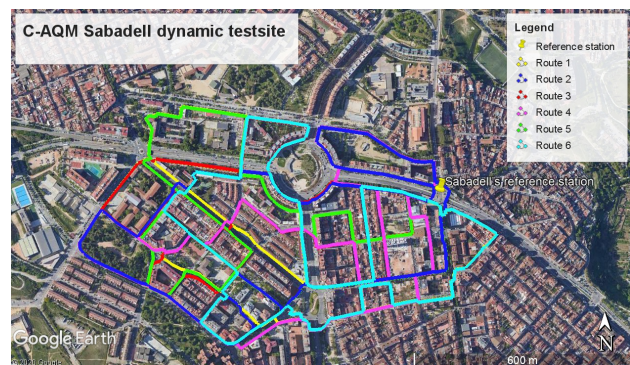


Figure 4. Walking routes.



Figure 5. Some volunteers after one acquisition campaign.

### 3.2 Results - system validation

For the validation of the system, a comparison was made between the value of the reference station and the C-AQM provided data. Table 2 shows the measurements of the reference station and data captured by the sensors in coordinates close to the station. The data obtained confirm the accuracy of the system and the results of the data capture campaigns.

| Day | Time | Num. of sensors |
|---|---|---|
| 11/21/2019 (Thursday) | 07:30 | 5 |
| 11/23/2019 (Saturday) | 07:30 | 5 |
| 12/12/2019 (Thursday) | 17:30 | 5 |
| 12/15/2019 (Sunday) | 17:30 | 5 |
| 03/05/2020 (Thursday) | 17:30 | 5 |
| 03/07/2020 (Saturday) | 08:00 | 5 |
| 09/21/2020 (Monday) | 07:30 | 4 |
| 09/22/2020 (Tuesday) | 07:30 | 4 |
| 09/28/2020 (Monday) | 17:30 | 4 |
| 09/30/2020 (Wednesday) | 17:30 | 5 |
| 10/05/2020 (Monday) | 17:00 | 5 |

Table 1. Dates of acquisition campaigns.

| Station/Sensors | Latitude | Longitude | $NO_2$ |
|---|---|---|---|
| Station | 41.561253 | 2.101017 | 27 |
| 1 | 41.560968 | 2.101198 | 27 |
| 2 | 41.561263 | 2.100958 | 28 |
| 3 | 41.560968 | 2.101199 | 30 |
| 4 | 41.560875 | 2.101213 | 26 |
| 5 | 41.560920 | 2.101213 | 28 |

Table 2. Comparison between C-AQM provided data and reference station (day 10/05/2020).

One of the criteria defined to generate the routes for data capture was that the sensors had to cross during the journey. This criterion enables the correct calibration of the measurements and confirms that the values captured by two sensors in the same place and at approximately the same time are similar (Table 3).

| Sensors | Time | Latitude | Longitude | $NO_2$ |
|---|---|---|---|---|
| 5 | 15:44:52 | 41.557585 | 2.094377 | 18 |
| 4 | 15:46:19 | 41.557585 | 2.094377 | 18 |
| 4 | 15:45:29 | 41.557752 | 2.095208 | 23 |
| 2 | 15:47:25 | 41.557752 | 2.095208 | 23 |
| 3 | 15:50:55 | 41.558548 | 2.092594 | 7 |
| 2 | 15:50:57 | 41.558548 | 2.092594 | 9 |
| 5 | 15:57:31 | 41.560525 | 2.094068 | 14 |
| 2 | 15:59:20 | 41.560525 | 2.094068 | 17 |
| 2 | 15:58:42 | 41.560161 | 2.093879 | 18 |
| 3 | 15:58:51 | 41.560161 | 2.093879 | 18 |
| 4 | 16:15:23 | 41.560877 | 2.101386 | 15 |
| 1 | 16:20:29 | 41.560877 | 2.101386 | 17 |

Table 3. Data from two sensors at the same coordinate and time (day 09/30/2020)

### 3.3 Visualization of results

In the QGIS Program, the calibrated data was used to estimate $NO_2$ levels where data capture was not performed within the study area. The interpolation method used in the project was Inverse Distance Weighted (IDW). In this method, the influence of a point with known values decreases as the distance increases, in this way, the influence of each point is proportional to the inverse of the distance.

The data visualization of each data capture campaign was carried out using 4 maps. One for the original data, the second for the data with corrected coordinates, the third for the final calibrated data, and the last for the interpolated data. The interpolated maps were presented to the volunteers for their feedback.

The second part of data visualization is done through the web viewer. The data defined for the viewer were: data processed and interpolated for every day of data capture, data from the official reference station for air quality measurement and as base maps, data from other WMS services (orthophoto, open street map and land use).

Using GeoServer, for each processed and interpolated data to be presented in the viewer, a WMS was created. The web viewer code uses the following programming languages: Hyper Text Markup Language (HTML), Cascading Styles Sheets (CSS), JavaScript and the OpenLayers library. Quickly and easily, using only a browser, the user has access to all the final results of the data capture campaigns, Figure 6.

Some of the features of the viewer are: download an informative guide to the viewer; load the coordinates and the point on the map, where the network where the user's computer is connected is located; select which layer and base map you want to display; click on the calibrated data points and obtain the $NO_2$ level at the respective point; download the map currently represented.
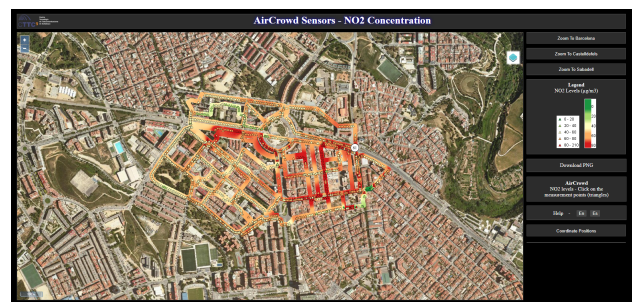


Figure 6. Web viewer.

### 3.4 Results - interpretation

Once the accuracy of the C-AQM system was validated, the data was analyzed. The comparison of maps of working days and weekends, shows a situation of lower pollution by $NO_2$. This fact was also observed by the participants who certified that the density of vehicles (the main source of this pollutant) was much smaller. Figure 7 (top) shows data for a working day, while bottom, the data is for a weekend of the same week. The captures were made at the same time, at 7:30 in the morning, and under approximately the same environmental conditions. It is observed that the $NO_2$ concentration on 11/21/2019 (Thursday) is higher than on 11/23/2019 (Saturday).

The comparison pre and post-pandemic on working days also shows slight improvement in the measures of air quality (Figure 8). The volunteers also declared that there was a decrease in road traffic before and after the first wave of the pandemic. Finally, the volunteers declared that the maps and the experience of collecting data have a strong impact on their perception of air pollution.

### 4. CONCLUSIONS

In this article, we have presented the first results of the pilot tests of the C-AQM system. The results obtained during the validation of the system confirm the accuracy of the C-AQM data after calibration and of the results obtained during the data capture campaigns in the city of Sabadell.

In the pre-pandemic data, a situation of lower $NO_2$ contamination is demonstrated on weekends in comparison with working days. Also related to the COVID-19 pandemic, analyzes of the data in the pre and post-pandemic period on working days also shows a slight improvement in air quality measures. These analyzes were also observed by the participants in the pilot test,
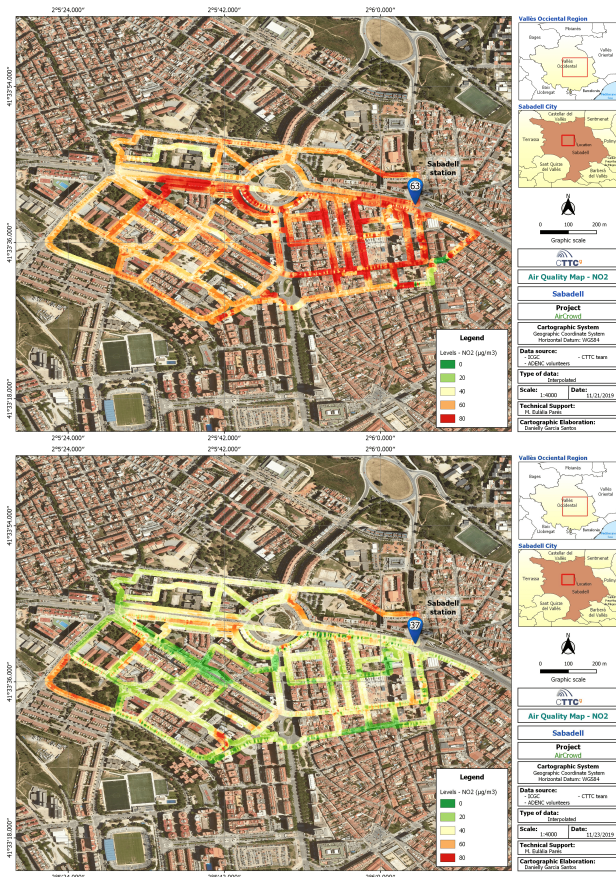
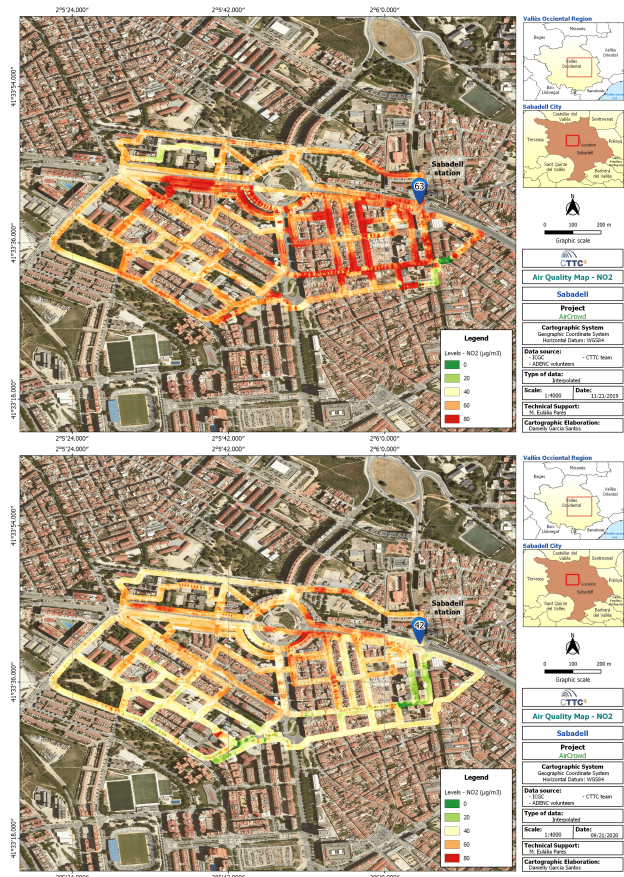Figure 7. Data set Top.- at 7:30 of a winter's Thursday and Bottom.- at 7:30 on a winters Saturday



Figure 8. Data set Top.- at 7:30 on a pre-pandemic Thursday and Bottom.- at 7:30 on a pre-pandemic Monday

who certified that the density of vehicles was lower on weekends and in the post-pandemic period.

The results of the project allow to conclude that the C-AQM system is a useful tool for citizen participation and learning about the problem of air quality in cities, as well as demonstrating its suitability for urban managers.

## ACKNOWLEDGEMENTS

## REFERENCES

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., Bartonova, A., 2017. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment international*, 99, 293–302.

ESA, 2017. Algorithm theoretical basis document for the TROPOMI L01b data processor. https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-TROPOMI-Level-1B-ATBD. Last accessed 17 April 2021.

Gerboles, M., Spinelle, L., A., B., 2017. Measuring air pollution with low-cost sensors. https://ec.europa.eu/jrc/en/publication/brochures-leaflets/measuring-air-pollution-low-cost-sensors. Last accessed 16 April 2021.

Jiménez, P., Jorba, O., Baldasano, J., Gassó, S., 2008. The Use of a Modelling System as a Tool for Air Quality Management: Annual High-Resolution Simulations and Evaluation. *The Science of Total Environment*, 390(2-3), 323–340.

Khomenko, S., Cirach, M., Pereira-Barboza, E., Mueller, N., Barrera-Gómez, J., Rojas-Rueda, D., de Hoogh, K., Hoek, G., Nieuwenhuijsen, M., 2021. Premature mortality due to air pollution in European cities: a health impact assessment. *The Lancet Planetary Health*. doi.org/10.1016/S2542-5196(20)30272-2.

Parés, M. E., Garcia, D., Vázquez-Gallego, F., 2020. Mapping air quality with a mobile crowd-sourced air quality monitoring system (C-AQM). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*,

XLIII-B4-2020, 685–690. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLIII-B4-2020/685/2020/.

Parés, M. E., Vázquez-Gallego, F., 2018. C-AQM: A crowd-sourced air quality monitoring system. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4, 491–497. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-4/491/2018/.