

# MAPPING AIR QUALITY WITH A MOBILE CROWDSOURCED AIR QUALITY MONITORING SYSTEM (C-AQM)

M. Eulàlia Parés\*, Danielly Garcia and Francisco Vázquez-Gallego

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

Commission IV, WG 10

**KEY WORDS:** Urban modelling, crowd-sourcing, air quality maps, smart cities.

## ABSTRACT:

World cities are currently facing one of the major crisis of the last century. Some preliminary studies on COVID-19 pandemia have shown that air pollutants may have a strong impact on virus effects. Improved gas sensors and wireless communication systems open the door to the design of new air monitoring systems based on citizen science to better monitor and communicate the air quality levels. In this paper, we present the Crowdsourced Air Quality Monitoring (C-AQM) system, which relies on Air Quality Monitoring reference stations and a cluster of new low-cost and low-energy sensor nodes, in order to improve the resolution of air quality maps. The data collected by the C-AQM system is stored in a time series database and is available both to city council managers for decision making and to citizens for informative purposes. In this paper, we present the main bases of the C-AQM system as well as the measurements validation campaign carried out.

## 1. INTRODUCTION

Air pollution is becoming one of the main threats for urban societies. Besides the laws and efforts to reduce the pollutants emissions, technicians and administrators are working hard to develop alert systems aiming to protect the more vulnerable citizens during high pollution episodes and to describe and recommend "green routes" for the most vulnerable citizens.

A key tool for that purpose are the reliable air pollution measurements. The Copernicus European system for Earth observation started providing measurements with high coverage and medium resolution (7x7km) around Europe (ESA, 2018). In addition, national and regional authorities are using maps generated from the interpolation of measurements acquired at reference stations located in relevant points of the territory (Jiménez et al., 2008).

Nowadays, the latest developments in air monitoring sensors and its price reduction open the door to the deployment of dense wireless networks of sensors, installed in buildings, which may provide high resolution and accurate air pollution maps in highly populated areas (Castell et al., 2017). Next step, in order to reduce costs and to improve performances is to be able to locate sensors in moving platforms. This approach allows to acquire data also in whole neighbourhoods but with a limited number of sensors.

In this paper, we present the validation process of the Crowdsourced Air Quality Monitoring (C-AQM) system (Parés, Vázquez-Gallego, 2018), which has been designed to generate high-resolution air quality maps using data from crowdsourced sensors and reference stations. As discussed in (Parés, Vázquez-Gallego, 2018) the development of the C-AQM system presents four main challenges: 1) to properly calibrate the low-cost sensors to ensure a performance good enough to use them as a complement of

reference stations and the Copernicus system; 2) the correct geo-referencing of all measurements, including those acquired inside urban canyons or tunnels; 3) the wireless transmission of measurements from thousands of sensor nodes to a Cloud server; and 4) the storage, analysis and representation of the information. In this paper, we shortly review the design and implementation of C-AQM and we mainly focus on the performance evaluation of the system regarding the calibration challenge. The C-AQM system has been deployed in the city of Sabadell, and six tests, one hours each, have been carried out. The results of these tests are presented in this paper.

## 2. SYSTEM DESCRIPTION

The architecture of the C-AQM system is based on three subsystems as shown in Figure 1: 1) acquisition subsystem; 2) processing subsystem; and 3) storage, analysis and visualization subsystem. The operation of the C-AQM system is briefly described as follows. Each user or data provider of C-AQM will install an AirCrowd device on his/her (non-motorized) vehicle. 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 and finally generate the results. The functionalities of each subsystem are described in detail in (Parés, Vázquez-Gallego, 2018). Hereafter, a brief description is done.

### 2.1 Acquisition 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)

\* Corresponding author

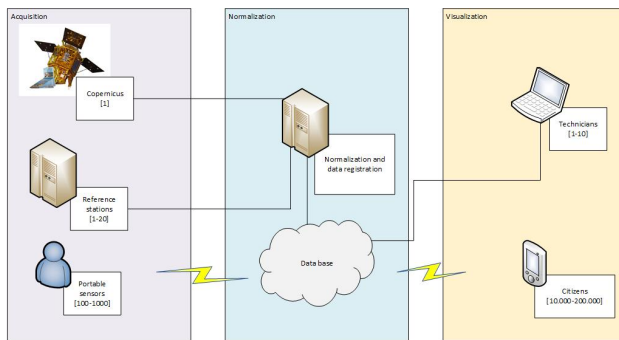


Figure 1. Architecture of the Crowd-sourced Air Quality Monitoring System.

official air quality monitoring reference stations, and 2) a network of hundreds or thousands of low-cost sensor nodes. In the next months, we aim to include also data from the Copernicus European system.

In our previous work (Parés, Vázquez-Gallego, 2018), we developed a low-cost sensor node, named AirCrowd (Air Quality Crowd-sourced sensing device), which implements 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 AirCrowd sensor node is composed of one CC2640R2 System-On-Chip from Texas Instruments, which integrates an ultra-low power microcontroller (Cortex-M3) and a Bluetooth Low Energy transceiver; one L80R GPS receiver from Quectel; one NO<sub>2</sub>, SO<sub>2</sub> and O<sub>3</sub> gas sensors from SPEC Sensors; one SM-PWM-01C particle sensor from Amphenol; and one BG96 wireless communications transceiver from Quectel. The BG96 is an ultra-low power consumption LTE Cat. M1 / Cat. NB1 (NB-IoT) / EGPRS module that offers a maximum data rate of 300 Kbps down-link and 375 Kbps up-link.

## 2.2 Processing Subsystem

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 AirCrowd sensor nodes (i.e., GNSS single frequency receiver) and also from public data like street maps information available at Open street maps. All data provided by these sources is introduced in a Differential Global Navigation Satellite System (DGNSS) / map-matching post-processing algorithm able to properly locate the measurements with one-meter accuracy, even in urban canyons or in short tunnels (Quddus et al., 2007).

The **calibration** process can be carried out thanks to the crowd-sourced nature of the C-AQM system. We can assume that some AirCrowd sensor nodes will acquire data near a reference station. Those that not, will at least acquire data in the same place, or near, where other sensor nodes did before. Thus, we can imagine the trajectories followed by all the C-AQM users as a

dense network with nodes that must share the same measurement values. We also assume that for short-term periods (less than 2 hours) the behaviour of sensor errors is quite stable and can be characterized mainly as a bias (Parés, Vázquez-Gallego, 2018). Consequently, a single least square adjustment of these network allows to calibrate all the sensors at once with an accuracy equivalent to the sensors noise.

## 2.3 Storage, Analysis and Visualization Subsystem

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<sup>1</sup> because it is reported as the current most popular time series database. With this selection we ensure not only that we are using a good enough solution, but also technical support from their developers in the mid-term.

The expert user interface relies on a Geographical Information System (GIS). The data stored at InfluxDB can be downloaded as a shape file (.shp). The expert user will be able to analyze, manage and represent the data in any GIS system like QGIS or ArcGIS.

## 3. PERFORMANCE EVALUATION

Hereafter the validation plan and main results for the communications subsystem and the Air Quality measurements subsystem are presented.

### 3.1 Communications Subsystem

The aim of these tests is to measure the performance of the communications subsystem in terms of time jitter. The tests have been done with 6 different AirCrowd sensor nodes that were programmed to acquire NO<sub>2</sub> concentration measurements and GPS positions every 6 seconds and transmit these data to a Cloud server through the NB-IoT cellular network of Vodafone-Spain.

The performance evaluation of the communications subsystem has been done in static and dynamic mode. In static mode, the AirCrowd sensor nodes are located in the same position for several hours. In dynamic mode, the AirCrowd sensor nodes are moved in three different scenarios as detailed below.

**3.1.1 Test Scenarios.** For the static mode tests, two different locations were selected as shown in figure 2. The first location is the roof of CTTC premises, which is located in a technological park in the city of Castelldefels, near the airport of Barcelona. The second location is the roof of a house in the center of Sabadell, a medium-size city of the metropolitan area of Barcelona. All acquisition were three hours length.

The dynamic mode tests were carried out in three areas as shown in Figure 3: inside the technological park where CTTC is located; along a 7 km segment of the road that connects the cities of Castelldefels and Viladecans; and finally, inside the city of Sabadell. The acquisition tests were 30 minutes length.

<sup>1</sup> <https://www.influxdata.com/>

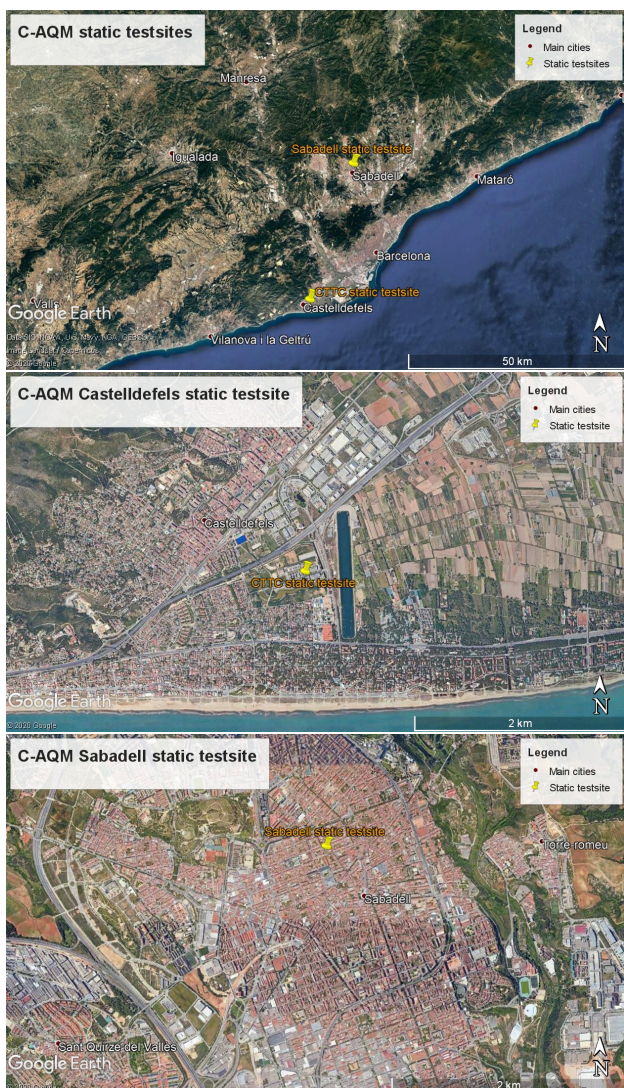


Figure 2. Location of static test-sites. Top.- Location within a global map; Middle.- Castelldefels test-site; Bottom.- Sabadell test-site.

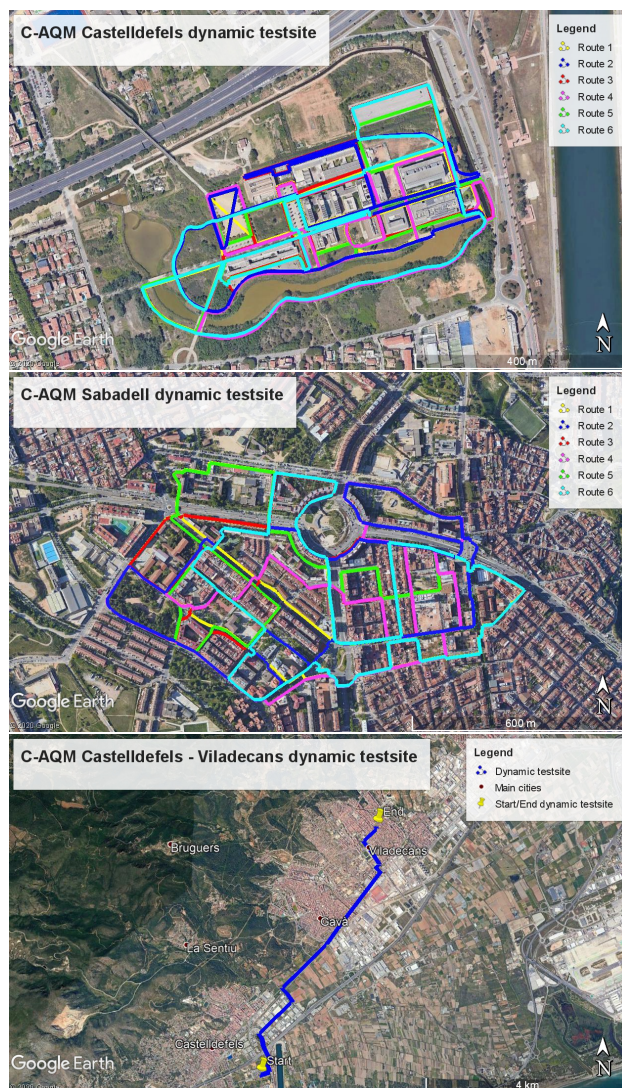


Figure 3. Location of dynamic test-sites. Top.- Castelldefels test-site; Middle.- Sabadell test-site; Bottom.- Castelldefels - Viladecans test-site.

**3.1.2 Tests Results.** In Table 1, the mean and standard deviation of the time period between consecutive data sets are presented. The statistics were computed for all the tests length. While the nominal time period is 6 seconds (i.e., data sets are periodically transmitted every 6 seconds by the sensor nodes), it can be observed that the system is mainly providing data sets every 6 seconds with a standard deviation between 1 to 5 seconds depending on the test mode and site.

| Testsite             | Performance |                          |
|----------------------|-------------|--------------------------|
|                      | Mean (s)    | Standard(s) deviation(s) |
| Static- Cast         | 6           | 2                        |
| Static - Sab         | 7           | 1                        |
| Dynamic - Cast       | 6           | 4                        |
| Dynamic - Sab        | 6           | 3                        |
| Dynamic - Cast/Vilad | 6           | 4*                       |

Table 1. Mean and standard deviation of time period between consecutive data sets

As it can be observed in Table 1, while in static mode the system is behaving as expected (i.e., the time jitter is very low), when sensor nodes are moving, the time jitter is too high. This fact may be due to the non-deterministic latency introduced by the NB-IoT network. This is specially relevant in the tests done between Castelldefels and Viladecans (\*). In the middle of the test, the connectivity was lost for a while and was recovered after several seconds, probably due to a hand-over between cellular base stations. In order to compute the standard deviation of that test, the connectivity loss time was not taken into account.

### 3.2 Air Quality Measurements: Comparison between Air-Crowd Sensor Nodes

The aim of these tests and the ones presented in next subsections is to evaluate and validate the performance of the C-AQM system as a tool for Air Quality Monitoring. Firstly, in this subsection we evaluate the behaviour of different sensor nodes under the same conditions of NO<sub>2</sub> concentration, temperature, relative humidity and pressure. More specifically, the objective is to compute the differences in the measurements of NO<sub>2</sub> concentration acquired by different sensor nodes as well as to

analyze how temperature and pressure changes affect the NO<sub>2</sub> concentration measurements.

**3.2.1 Test Scenarios.** For the first tests set, six sensors were placed at CTC premises roof at same time and under same environmental conditions (sun, wind, atmospheric pressure and humidity) during at least four hours. Tests were carried out in different conditions: high ad low pressures, sunny and cloudy days, foggy and dry days, short and long acquisitions. Second tests was carried out in Sabadell, moving sensors together through the same path, also in different atmospheric conditions (windy, dry, sunny,...)

**3.2.2 Tests Results.** For each test, a sensor was selected as the "pivot" one and the rest were compared against it. The difference of measurements when acquiring data at the same time is computed. In table 2, the mean values of those differences and its standard deviation can be checked. The mean values differ from -60 to 60 (ug/m<sup>3</sup>), depending on the environmental conditions, but what is more relevant is that the standard deviation is quite similar, around 10 (ug/m<sup>3</sup>). This means that all the sensors have the same behaviour and follow the same trends.

| Test type | Duration (hour) | NO <sub>2</sub> difference between 2 sensors |   |
|-----------|-----------------|--|---|
|           |                 | Mean (ug/m <sup>3</sup> )                    | Standard deviation (ug/m <sup>3</sup> ) |
| Statics   | 4               | -60 to +60                                   | 9 ± 4                                   |
| Dynamics  | 0.5             | -70 to +70                                   | 12 ± 3                                  |

Table 2. NO<sub>2</sub> concentration differences of AirCrowd sensor nodes

### 3.3 Air Quality Measurements: Comparison between AirCrowd Sensor Nodes and Reference Stations.

A second set of tests were done in order to check that effectively, in short-term periods (<2 hours) the systematic error can be modeled as a bias and the non-systematic errors are small enough to provide acceptable data. Last set of tests check that the calibration SW is able to properly estimate and remove the systematic errors of the measurements.

**3.3.1 Test sites.** Second test was carried out near a the Viladecans reference station owned by Generalitat de Catalunya (see Figure 4), which data is available on the web (Departament de Territori i Sostenibilitat, 2020). The tests have been realized in static at 25°C for 1.5 hours (short static) and 20 hours (long static).

**3.3.2 Main results.** The errors of the NO<sub>2</sub> concentration measurements in short static and long static tests are shown in Table 3. Figure 5 shows the measurements of NO<sub>2</sub> concentration acquired by a reference station and the NO<sub>2</sub> sensor of the AirCrowd sensor node within a time period of 20 hours. As it can be observed in Table 3, the analysis of the NO<sub>2</sub> concentration measurements shows, in one hand, a relevant bias in both short static and long static tests. However, the drift of the bias is small enough to consider the bias as constant within a temporal window between two and three hours. On the other hand, as it can be observed in Figure 5, the NO<sub>2</sub> measurements are rather noisy, with a standard deviation of around 7 μg/m<sup>3</sup>, mainly due to a large quantification error of the NO<sub>2</sub> sensor. As it can be observed in Figure 5, by applying a moving average filter of 1 hour and a correction of a constant bias on the measurements provided the NO<sub>2</sub> sensor of the AirCrowd sensor node,

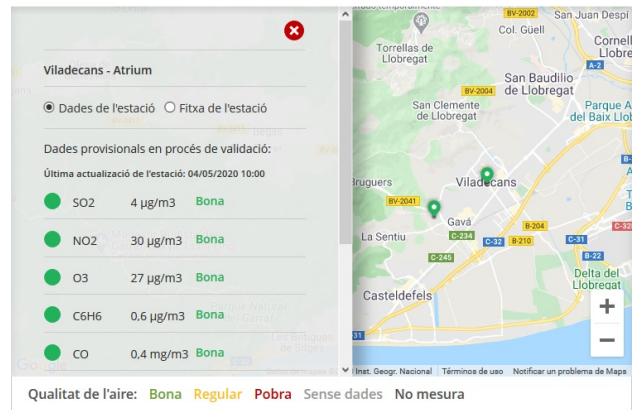


Figure 4. Viladecans reference station from DTES.

the results achieved fulfill the precision and accuracy requirements (better than 25%) defined in (Parés, Vázquez-Gallego, 2018).

| Test type    | Duration (hour) | NO <sub>2</sub> sensor     |                           |
|--------------|-----------------|----------------------------|---------------------------|
|              |                 | Noise (ug/m <sup>3</sup> ) | Bias (ug/m <sup>3</sup> ) |
| Short static | 1.5             | 7.6                        | 6.0-42.0                  |
| Long static  | 20              | 7.2                        | 23.0                      |

Table 3. NO<sub>2</sub> concentration errors of the AirCrowd sensor node

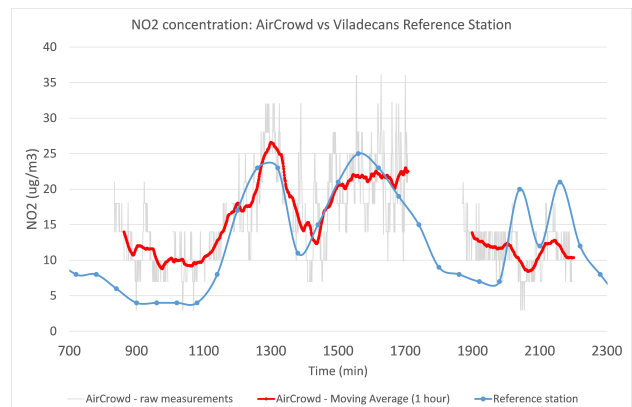


Figure 5. NO<sub>2</sub> concentration: AirCrowd sensor node vs Reference station measurements along 20 hours

### 3.4 Air quality measurements system performance: Calibration procedure evaluation.

Last set of tests check that the calibration SW is able to properly estimate and remove the systematic errors of the measurements.

**3.4.1 Test sites.** These tests were carried out in the city of Sabadell, all tests were carried out having as starting and ending point the Sabadell's reference station owned by Generalitat de Catalunya (see figure 3-middle, which data is available on the web (Departament de Territori i Sostenibilitat, 2020). Three different campaigns were done. Each of them consist on collecting data at the same hour (7:30 morning or 17:30 afternoon) during one hour in two different days (mainly Thursday and Saturday/Sunday). The sensors were carried out by volunteers that walk through the city with the sensors in a net handbag. In each campaign, measurements of 5 sensors were collected.

**3.4.2 Main results.** As it can be seen in figure 6, the measurements of the different sensors can be clearly differentiated. The values are quite similar during each of its paths but completely different from one to another sensor. Once the least square adjustment is done, including the value of the reference station, located in one of the borders of the test site, all the sensor measurements improves significantly and a plot closer to reality can be seen. Since in those tests we can not have check points other than the reference station, we use the reference station values when the volunteers visited it for second or third times as check points. The computed vs real values of those checkpoints can be seen in table 4. Please, note that at the moment being just three tests for each class (working day / weekends) has been done. We expect to do more campaign and the presented figures to be improved in the near future.



Figure 6. Data set Top.- Before and Bottom.- after calibration process

| Tests type   | Duration (min) | NO2 difference C-AQM / ref. station Difference (mean) (ug/m3) |
|--------------|----------------|---|
| Working days | 45             | 12  |
| Weekends     | 45             | 8   |

Table 4. NO2 estimated error at checkpoints.

#### 4. C-AQM APPLICATIONS

The analysis of the first experiences with C-AQM let us to discard, for the moment, the use of C-AQM as a tool for providing

official absolute measurement of air pollutants to the administration. However, C-AQM is proving to be an excellent tool for i) urban planners, ii) social services and iii) educational services as a pollutant map service. With a system like C-AQM, measuring not only NO2 but also other pollutants, urban planners can have a better idea of suitability or unsuitability locations for critical services. Hot spots can be detected of the analysis of several maps of the same neighborhood at different times. (See, for example, 7). Social services will have the chance to better define "green paths" for vulnerable people to go from a place A to a place B, and last but not least, since one image is more powerful than a thousand words, having pollutant maps at street level, can help the citizens to be more aware of the quality of the air they breathe.

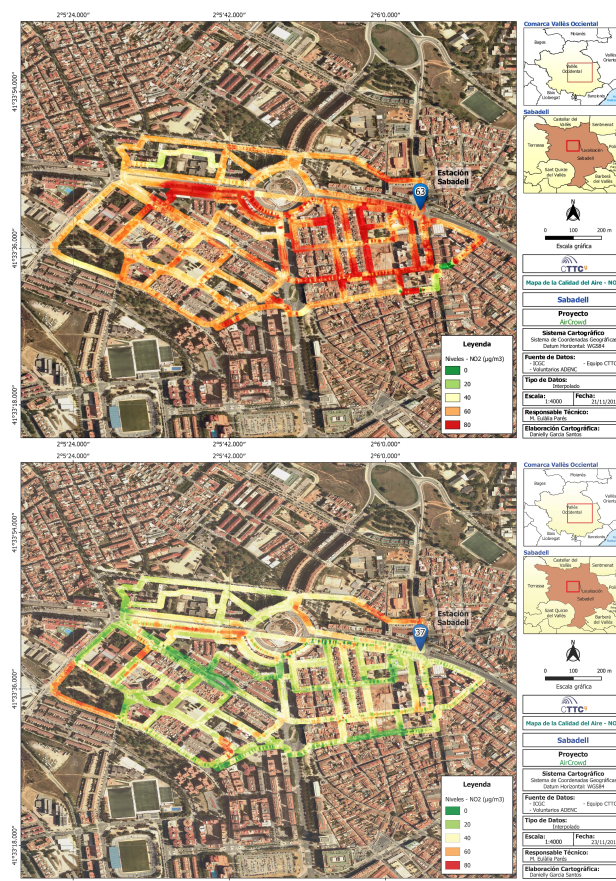


Figure 7. Data set Top.- at 7:30 of a winter's Thursday and Bottom.- at 7:30 on a winters Saturday

#### 5. CONCLUSIONS AND FURTHER WORK

In this paper, we have presented the idea behind the Crowdsourced Air Quality Monitoring system (C-AQM) and the first results of the validation campaign. Based on the current state-of-the-art and the available technologies, we have designed a system that combines air quality measurements obtained from the available reference stations, and a cluster of low-cost low-energy sensor nodes. By jointly processing these measurements, the system is able to generate high resolution (25x25m) air quality maps. The initial validation exercises allow us to be optimistic on the suitability of the system for urban managers. The short-term stability of the sensors (bias stable within 60

minutes) is good enough to be used in a system that should allow sensor re-calibration every few minutes (20 minutes maximum between calibrations), and the GNSS/map matching technology provides enough accuracy (below 2 meters) for the application. Despite of the promising preliminary results, the project is still on its initial phases and more sets of dynamic tests under several environmental conditions is still needed.

### ACKNOWLEDGEMENTS

The authors would like to acknowledge Sabadell city council and their program "Pressupostos Participatius 2019" for partially funding this work. We would also like to thank AD-ENC (Associació per a la Defensa i l'Estudi de la Natura de Catalunya) for their support to the project and their unvaluable help in testing site volunteers recruitment. the project is supported by Generalitat de Catalunya under the program Suport als Grups de Recerca Emergents (2017 SGR 820). Last but not least, the authors also would like to thank all the volunteers for the walks and the commitment with the project.

### 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.
- Departament de Territori i Sostenibilitat, 2020. <http://mediambient.gencat.cat/ca/05ambitsdactuacio/atmosfera/qualitatdelaire/vols-saber-que-respires/>.
- ESA, 2018. Sentinel 5. <https://earth.esa.int/web/guest/missions/esa-future-missions/sentinel-5> (28 May 2018).
- 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.
- Parés, M. E., Vázquez-Gallego, F., 2018. C-AQM: A crowd-sourced air quality monitoring system. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-4, 491-497. doi.org/10.5194/isprs-archives-XLII-4-491-2018.
- Quddus, M. A., Ochieng, W. Y., Noland, R. B., 2007. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation research part c: Emerging technologies*, 15(5), 312–328.