SenseBoard: Sensor Monitoring for Air Quality Experts

Federica Rollo
'Enzo Ferrari' Engineering Department
Modena, Italy
federica.rollo@unimore.it

Laura Po
'Enzo Ferrari' Engineering Department
Modena, Italy
laura.po@unimore.it

ABSTRACT

Air quality monitoring is crucial within cities since air pollution is one of the main causes of premature death in Europe. However, performing trustworthy monitoring of urban air quality is not a simple process. Especially, if you want to try to create extensive and timely monitoring of the entire urban area using low-cost sensors.

In order to collect reliable measurements from low-cost sensors, a lot of work is required from environmental experts who deploy and maintain the air quality network, and daily calibrate, control, and clean up the data generated by these sensors. In this paper, we describe SenseBoard, an interactive dashboard created to support environmental experts in the sensor network control, management of sensor data calibration, and anomaly detection.

1 INTRODUCTION

Air pollution is a global threat leading to large impacts on human health and ecosystems, particularly in urban areas. In Europe, air quality remains poor in many cities that experience exceedances of the regulated limits for air pollutants [1]. The urgency of limit air pollution is also stated by the sustainable development goals (SDGs) defined in the 2030 Agenda for Sustainable Development [2].

Effective action to reduce air pollution and its impact on the quality of life requires good understanding and extensive monitoring of urban air quality. In recent years, the development of Internet of Things technologies has increased and cities around the world have exploited this enabling technology to be able to control multiple aspects of citizens' lives. IoT allows monitoring traffic congestion [3, 10], detecting and classifying road accidents [7], managing car parking [9], supporting decision in agriculture [4], evaluating energy consumption [12], and, also, monitoring air quality [13]. Data generated by IoT are used to improve city services and the living experience of citizens.

In this context, data coming from a group of low-cost sensors spread around a city might generate widespread hyperlocal insights into air pollution. However, a network of low-cost air quality sensors is not enough to monitor urban air quality. Since those sensors are complex and sensitive, they require specific environmental skills. Data generated by the air quality sensors need to be converted into relevant and crucial insights to allow the monitoring of air quality by politicians and to enable the achievement of the sustainability goals. In this context, environmental experts hunger for a control platform to perform sensor data calibration and anomaly detection.

The maintenance and control of a urban air quality network is relevant, and crucial to provide good information that enables the extensive monitoring of air quality. Moreover, the availability of effective visualizations to process and interpret collected data is essential.

In this paper, we present SenseBoard, an interactive tool addressed to environmental experts that brings together heterogeneous and dynamic data for real time analysis and management of air quality network. This tool has been conceived within the TRAFAIR project[1] that allowed the creation of an urban network of low cost air quality sensors. The low-cost sensors employed are cheaper and less reliable than the Air Quality Monitoring (AQM) legal stations managed by the Environmental Agencies. It is possible to improve the reliability of the measurements of these devices if they are previously calibrated by placing the device near air quality stations for some weeks. Low-cost sensors provide "raw" measures, i.e. a datum in millivolts; to convert this datum into a reliable concentration of pollutant it is necessary to carry out a calibration period during which some Machine Learning algorithms are trained in order to generate, from the raw measurements, pollutant concentrations in line with those estimated by the AQM stations. SenseBoard is devoted to support environmental experts in the monitoring and control of the air quality sensor network, in the supervision of the calibration process and in the detection of anomalous values. SenseBoard acts as an enabling tool to detect anomalies, update sensor status, monitor the proper functioning of the sensors, manage the change of location of the devices and, above all, to provide feedback to perform the calibration process. The calibration results obtained using the Machine Learning algorithm are shown and compared to the raw data, and the data of the AQM station, thus, it is possible to understand if the algorithm works appropriately or if it is necessary to extend the co-location period of the device.

SenseBoard is a general and flexible dashboard that can be adapted for the monitoring of any air quality sensor network. The scalability of the dashboard allows replicability in cities of different size with a variable number of sensors. The dashboard is not affected by the type of employed sensors and it can be easily modified to visualize other parameters measured by the sensors. In this paper, we take advantage of the use case in the city of Modena.

The rest of the paper is organised as follows. Section 2 is devoted to the presentation of the background. Then, Section 3 introduces the dashboard and describes some views (data visualization) in the city of Modena. In the end, Section 5 provides conclusions.

2 BACKGROUND

TRAFAIR ("Understanding Traffic Flows to Improve Air Quality") [11] is a project co-financed by European Commission that brings together 10 partners from two European countries (Italy and Spain) to develop innovative and sustainable services combining air quality, weather conditions, and traffic flows data. The scope is to increase the awareness on urban air quality for the benefit of citizens and government decision-makers. The project aims

[1] https://trafair.eu

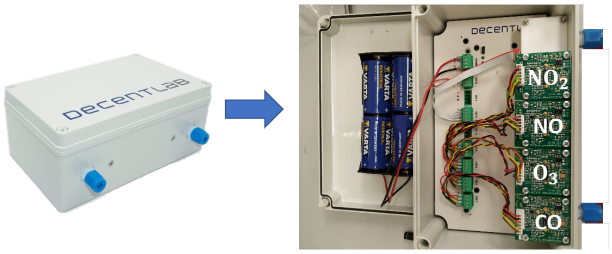

Figure 1: An air quality device (on the left) and its content inside (on the right): 4 cells/sensors for measuring the level of 4 air pollutants (NO, NO_2, CO, and O_3 in this case).

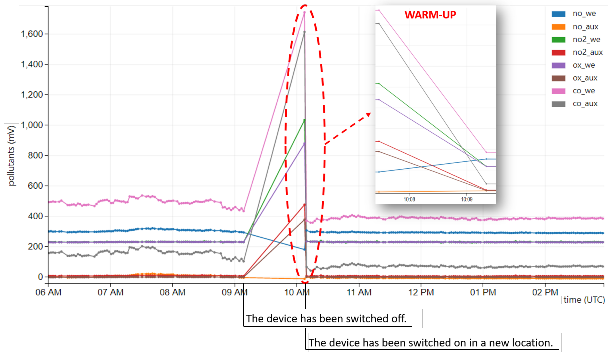

Figure 2: Raw observations of one air quality device made in two different locations.

to supervise the level of pollution on urban scale in 6 European cities (Modena, Florence, Pisa and Livorno in Italy, and Santiago de Compostela and Zaragoza in Spain) by producing real-time estimates of air pollution through a network of air quality devices and by developing a service for forecasting urban air quality based on weather forecasts and traffic flows [5].

The sensors employed are low-cost, cheaper and less reliable than the AQM stations managed by the Environmental Agencies. However, these devices can provide reliable measurements if they are co-located for a certain period of time close to the stations where they "learn" how to measure air quality.

A device is a box where different sensors are placed. Each sensor, also called cell, is devoted to the measurement of a specific pollutant. Figure 1 shows an exemplar device.

The approach used in TRAFAIR is to install the devices for a certain period close to the AQM stations. During this period (calibration period) a Machine Learning algorithm is trained on the measurments provided in millivolts by our devices (raw observations) and the AQM station measurements. Then, when environmental experts evaluate the device is "ready" (usually after 3 weeks of co-location), it can be moved to different location and start providing air quality measurements. When the device is "ready", thanks to the Machine Learning algorithm previously trained, calibrated data (concentrations) are generated from the raw observations. Usually, periodically every 6 months, the devices are once again co-located near the AQM stations for re-calibrating, thus maintaining a good quality of measurements.

In this scenario, it is easy to understand the importance of having a tool for managing the change of location or status of the devices. Besides, it is important to compare anytime the calibrated data with the measurements of the AQM stations (no matter in which location the sensor is), to determine when the device needs to be re-calibrated (usually when the calibrated data and AQM measurements differ a lot).

Since devices are constantly moved, it is possible that, when they are switched off and then switched on in a new location, they experienced a "warm-up" for some minutes or even hours. The warm-up is a specific period when the device tries to achieve a thermo-mechanical balance in the measuring system as well as an optimal operating temperature of the electronic components. Since the warm-up period is of variable duration, a time line evolution of the raw observations enables the user to detect when the warm-up period is over.

Figure 2 shows the raw measurements collected from one device. As it can be seen, at 9 a.m. approximately the device has been switched off. One hour later, the device has been switched on in a new location. The zoomed area of the graph shows the first two measurements for each gas made in the new location. The values drop off sharply due to the warm-up. Consequently, these values have to be discarded from the reliable observations and flagged by the environmental experts as "not reliable". This operation can be made through SenseBoard (this will be described in Section 3). Accordingly, the environmental experts need to constantly visualize the data produced by the device and monitor the behavior of each device.

The need for a monitoring tool comes from the environmental experts. For this reason, they have been directly involved in the definition of the requirements to be satisfied through the dashboard. After some discussions, 6 requirements have been outlined:

- **R1** providing an overview of the current position and status of each sensor,
- **R2** recording the location or status change for a certain sensor without hand-writing the SQL query to store the modification on the database,
- **R3** visualizing sensor observations without data aggregation to detect anomalous values and compare them each other, and with data aggregation to better understand the trend,
- **R4** comparing observations of co-located sensors in the same place,
- **R5** showing the concentrations produced by the calibration algorithm and comparing them with the certified values of the AQM legal stations,
- **R6** displaying the anomalies identified by the anomaly detection algorithm to control the efficiency of the automated algorithm.

After the development of SenseBoard, during the usage, additional feedback from environmental experts has been continuously collected to improve the functionalities of the dashboard.

2.1 Air quality sensor network

In Modena, an Italian city of 186,000 inhabitants and 183 Km^2, 13 air quality devices (52 low-cost sensors) have been installed in different locations. There have been identified 2 locations for calibration (the red dots in Figure 3), close to the AQM stations, and 10 locations of interest (the blue dots in Figure 3), that are placed in areas of different kinds, such as residential, industrial, or green areas.

Two types of low-cost devices have been exploited: 12 Decentlab Air Cubes and 1 Libelium Smart Environment PRO. All the devices are equipped with 4 cells (sensors), one for each gas (NO, NO_2, CO and O_x). Each cell measures the gas concentration

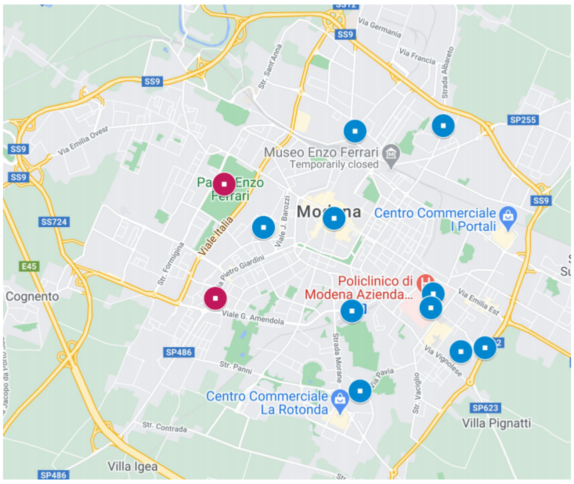

Figure 3: Points of interest for air quality monitoring (blue dots) and positions of the AQM stations (red dots). Map data: Google, 2020.

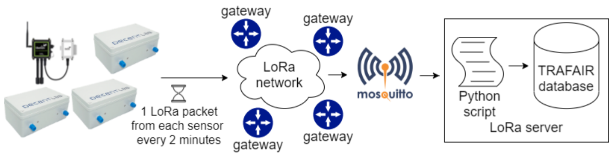

Figure 4: Sensor data acquisition from the low-cost air quality sensor network.

through 2 channels (the auxiliary and the working channels). In addition, the Libelium device measures the level of $PM_{2.5}$ and PM_{10}. For each channel, the raw observations are provided in mV, moreover, a basic concentration based on the original factory calibration[2] is provided in $\mu g/m^3$. Besides, these devices are able to measure the air temperature and humidity, and provide the battery voltage. Therefore, the total number of measurements provided by one sensor is 19 for Decentlab cubes and 21 for Libellium sensor.

The sensor data acquisition is managed by the Long Range Wide Area Network (LoRaWAN) implemented in the city of Modena. LoRaWAN [6] is a media access control (MAC) protocol widely used in smart city applications thanks to its easy installation and cost-effectiveness. It employs some gateways i.e. antennas that receive broadcast messages from the enabled devices (the air quality devices, in our use case) and forward them to the server. The message from one device can be received by more gateways at the same time, the server will deal with duplicates. The LoRaWAN network exploits low radio frequencies and provides for long-range communications (up to five kilometers in urban areas, and up to 15 kilometers or more in rural areas). The network coverage depends a lot on the geographic landscape.

Our air quality devices have been registered to the LoRaWAN network of Modena through their identifier (DevEUI) and following the Over-the-Air Activation (OTAA) process. The data rate has been set up to 125 kHz, and the spreading factor to 7, to allow devices transmitting data every 2 minutes.

[2]This is obtained by applying to the raw observations a formula provided by the manufacturing company with the calibration parameters that are different for each device.

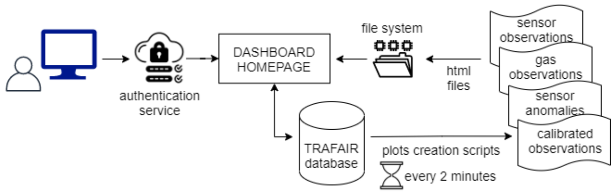

Figure 5: SenseBoard architecture.

Figure 4 shows the data acquisition process. The devices encapsulate the acquired measurements into LoRa packets and send them to the LoRaWAN server through the gateways. Every LoRa packet contains 19 (or 21) measurements (both mVs and basic concentrations for each gas and channel, including air temperature, humidity and battery voltage) and in one day each device produces 13,680 (19 * 720) measurements. 4 gateways have been installed in Modena, mainly on the roofs of the highest buildings, to cover the whole urban area of the city and ensure the coverage of the LoRaWAN network in our points of interest. Moreover, the gateways have been registered on the LoRa server. When the gateways receive a message, they send it to the LoRa server where the MQTT (Message Queue Telemetry Transport) Broker Mosquitto [8] is running. This is a publish/subscribe messaging transport protocol. Data are published as an MQTT topic. We have used the paho-mqtt Python library[3] to implement the open source message broker in a Python script. The script is always running and exploits the client class to enable the connection to the MQTT broker, publish messages, subscribe to topics and receive messages. Then, messages are decoded and measurements stored into a PostgreSQL database, the TRAFAIR database (in the following sub-section, more details are provided). Also, through this script an anomaly detection algorithm is applied to the time series of the air quality measurements to detect if each measurement is anomalous or not. This algorithm employs a majority voting system of three different Machine Learning algorithms. The anomalous data are flagged into the TRAFAIR database. When a device is moved from one location to another, it automatically connects to the nearest gateway and restart sending messages. Since the messages received in the LoRa Server are described by the identifier of the device, the change of gateway to which the device connects is completely transparent. The LoRa server keeps storing the measurements of each device no matter how they are moving in the urban context.

2.2 Data platform

Data from the air quality low-cost sensors are stored, in real time, into the TRAFAIR database. This database exploits the PostGIS extension to handle with geospatial data and the Timescale extension to make SQL scalable for time-series data.

The database contains more than 60 tables and 190 GB of data collected from the beginning of the TRAFAIR project (November 2018) till now (February 2021). Air quality measurements and device-related information are store in 11 tables and take 3 GB. These tables stores the technical characteristics of each device, its position, its status (running, calibration, offline, broken, warm-up), the raw observations, the concentrations obtained by both the original factory calibration and our calibration algorithm, and the anomalies identified by some anomaly detection algorithms applied to both raw and calibrated observations.

[3]https://pypi.org/project/paho-mqtt/

In each moment, every device is described by a status and is located in a point of interest (see Figure 3). Its measurements are stored continuously, as soon as they are parsed by the LoRa server. Each raw measurement can be calibrated by multiple calibration algorithms. Thus, calibrated data are identified, not only by the date of the measurement and the sensor that has provided it, but also by the algorithm that was used. In the end, several anomaly detection algorithms are applied to both raw and calibrated data. The results are stored in the TRAFAIR database in appropriate tables using boolean values to indicate if they are anomalous or not.

Only considering the measurements coming from our devices, from the installation, we have collected 3.3 million records of measurements (1.8 GB). Each record includes 19/21 measurements: air temperature, humidity, battery voltage, 8 raw measurements (2 channels per 4 gases), 8/10 concentrations of the original factory calibration (2 channels per 4 gases and one measure for $PM_{2.5}$ and PM_{10}).

3 SENSEBOARD

SenseBoard[4] is a Python web application which exploits Tornado[5] as web framework. It runs on a Debian 9 machine with 32 Intel(R) Xeon(R) Silver 4108 CPU at 1.80GHz processors and 256 GB RAM. Figure 5 shows the architecture of the dashboard. Firstly, users need to login to access the dashboard. The authentication phase is performed through the Lightweight Directory Access Protocol (LDAP). The list of people allowed to access is currently limited to the environmental experts working in TRAFAIR.

After the authentication, the user is able to visualize the current status of each device and send other requests through the navigation bar at the top: he/she can ask for observations (raw measurements), anomalies, calibration (calibrated measurements), and AQM station (measurements from the AQM stations). For each request, the dashboard queries the TRAFAIR database to obtain the appropriate data and creates plots of the time series data by using the matplotlib Python library[6]. More complex plots are periodically generated by ad-hoc Python scripts[7] which query the database and save plots in the file system as html files through the save_html function of the mpld3 library[8]. This library is also exploited for the InteractiveLegendPlugin[9], which allows to connect the plot to an interactive legend. This legend is very useful in our plots since it allows customizing the visualization by adding or removing some lines in the plot. The user can click on the rectangle generated in the legend near the labels. If the rectangle is colored, the corresponding data is shown on the plot; if the rectangle is white, these data are removed from the plot. The html files are, then, included in the html page of the corresponding request. What we mean with "more complex plots" are the ones which require an elaboration of the data stored in the database and manage a big amount of data (i.e. the raw observations of each sensor related to one month). This choice was made to save time in the visualization of the plots. Indeed, this solution decreases the server response time of 35 seconds for the most time-consuming request.

[4]https://trafair-srv.ing.unimo.it/aqsensors
[5]https://www.tornadoweb.org/
[6]https://matplotlib.org/
[7]The scripts run every 2 minutes and generate the plots in 4-27 seconds.
[8]https://mpld3.github.io/
[9]https://mpld3.github.io/examples/interactive_legend.html

Some examples of visualizations (views) are described in Section 3.2. The views are static to allow users to navigate and explore all the plots in the view without any interference. However, the user can click on the "update" button to see the updated views.

3.1 Users and scope

The scope of SenseBoard is the monitoring and control of the air quality sensor network and the supervise of the calibration and anomaly detection processes.

Regarding the monitoring of the network, SenseBoard allows to identify and update the status of the sensors, change their location when they are moved in different position and perform any maintenance, if necessary.

Considering the supervise of the calibration process, SenseBoard lets to compare raw measurements of co-located sensors, raw and calibrated measurements of the sensors, and, in particular, the calibrated observations generated during the calibration period with the legal observations from the AQM stations. This last operation is the crucial one in the calibration process because it allows experts to understand if the training period of the Machine Learning algorithm is sufficient, i.e. if the concentrations, elaborated by the Machine Learning algorithm, are in line with those of the AQM station.

Other tasks are the detection of issues in the network communication, the discovery of disruptions or failures in the sensor's behaviour, the identification of anomalous gas concentrations, the comparison of co-located sensors measurements, the correlation study of the pollution level in the area of sensor installation.

The primary users of our visual analytic dashboard are the environmental experts in charge of installation, maintenance and calibration of air quality sensors.

3.2 Views

In SenseBoard, we have developed 6 views to allow environmental experts to have complete control of the air quality sensor network status and the operations that are performed on the sensor data. Each view is described in detail in the following sub-sections.

3.2.1 *Sensor status and position*

The first view, i.e. the homepage of the dashboard after the login, aims at satisfying requirements R1 and R2. Here, users are able to visualize a table with a summary of the main information related to the air quality devices. For each device, in the table, there are listed its identifier, the name of the location where the device is currently installed, the timestamp of the installation, the name of the person in charge of the installation, the sensor status and any possible notes.

Besides, as shown in Figure 6, for each device, two buttons are available: the "edit" button allows to update the location and/or the status of the corresponding device. After clicking on the button, the user has to specify the timestamp representing the instant of the update (of the location or status), the location (one of the points of interest in Figure 3), the status, and, optionally, its name and notes. The "save" button stores the information in the TRAFAIR database. The status update is exploited in different situations. For example, if the device is moved from a point of interest to the AQM station, its status changes from "running" to "calibration". In addition, if the environmental experts notices an abnormal behavior of the device, he/she can modify the status in

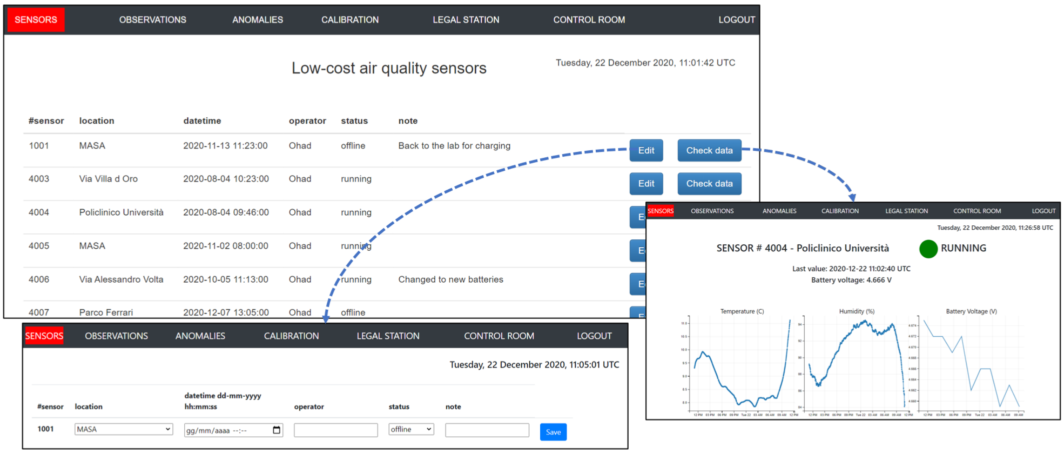

Figure 6: “Sensor status and position” view.

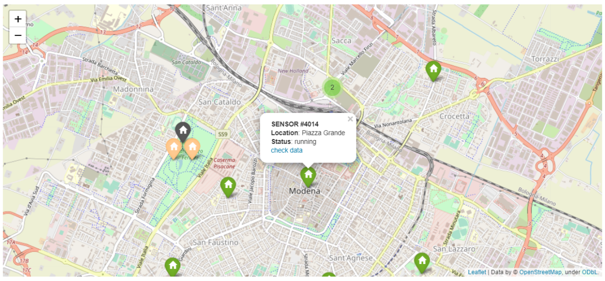

Figure 7: Position of the devices on January 4^{th}, 2021.

“broken” indicating as timestamp the date of the first abnormal measurement. Then, he/she needs to add the “running” status from the first regular measurement.

In addition to the “edit” button, the “check data” button connects to the “sensor observations” view.

Besides, in this view, the user can interact with a map (Figure 7), where the current position of each device is visualized with an icon of different colors according to the status of the device. If more devices are in the same location, a bigger icon is displayed on the map with the number of devices in that position. By clicking on this icon, an icon for each device is visualized. If you click on the icon of a device, you can see its name, its status, the name of the location, and the link to the “sensor observations” view of that specific sensor. Folium[10] is the Python library used to create the map.

3.2.2 *Sensor observations*

The “sensor observations” view satisfies requirement R3 and includes 6 plots with the observations of one device. At the top of the page, the name of the device, its status and location, and the timestamp of the last observation with the level of battery voltage are reported. This allows the managers of the sensor network to check immediately if the sensor is not sending data or if the batteries need to be changed.

The 6 plots show the measurements of:

(1) the relative humidity in percentage (%),
(2) the temperature in Celsius degree,
(3) the battery voltage in Volt (V),
(4) the raw observations of the 4 gases for auxiliary and working channels in mV,
(5) the observations of the 4 gases for auxiliary and working channels calibrated through the original factory calibration in $\mu g/m^3$,
(6) the observations calibrated through the TRAFAIR calibration algorithms in $\mu g/m^3$.

Only for the Libelium device another plot is provided, which shows the level of $PM_{2.5}$ and PM_{10}.

Each plot can visualize data for different time interval (last 24 hours, week, or month) and data aggregation (2 minutes - which means no aggregation, 5 minutes, and 15 minutes), generating 9 different combinations for each plot. The visualization changes according to the option selected by the user.

There are altogether 711 plots (13 devices * 6 plots * 3 time interval * 3 data aggregation + 1 PM plot * 3 time interval * 3 data aggregation). Since the creation of a plot took on average 12 seconds, we decided to generate these plots asynchronously through one Python scripts. This means that the plots are generated independently by the user choice, and when the user selects an option (for time interval and data aggregation), he/she enables the visualization of a ready-made plot. This time-saving design choice is also motivated by the user behavior. After three months from the first release of SenseBoard, we noticed that it was very likely that the user is interested in visualizing several plots, exploring different gases with different aggregations or for a different time interval. If the plots are created synchronously with the user’s choice, jumping from one plot to another requires waiting for the generation of the relative plot each time. In agreement with environmental experts, we have therefore decided to switch to an asynchronous generation of the plots that reload the 711 plots every 2 minutes.

Figures 8 and 9 are two examples of visualization available in the “sensor observations” view. In Figure 8 the measurements of the 4 gases for the auxiliary and working channels related to one device are plotted in a lines chart. An anomalous behavior of the device has been highlighted in red: the values of the measurements in that time interval are very different from the previous ones. SenseBoard allows the detection of the wrong data. After the maintenance work by the environmental experts, the device reaches the stability and the measurements proceed with the expected values. Through the “edit” button of the “sensor status and position” view, the time period related to the red

[10]https://pypi.org/project/folium/0.1.5/

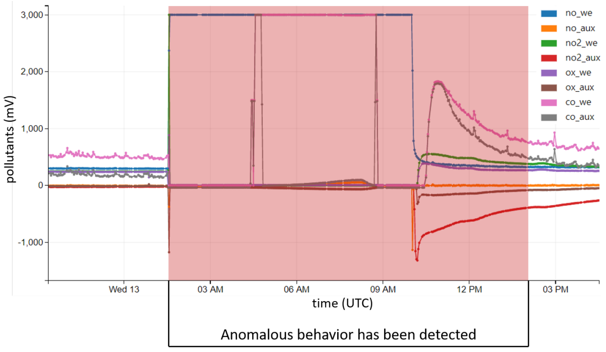

Figure 8: An anomalous behavior of a device detected on January 13^{th}, 2021.

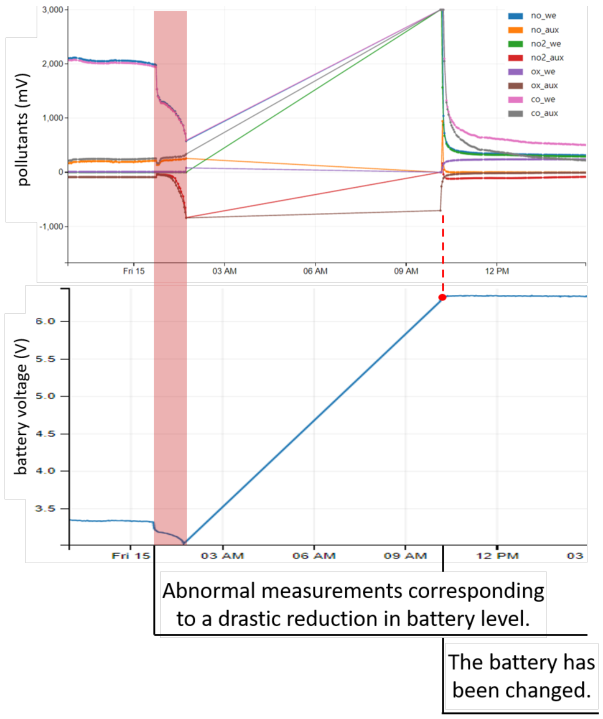

Figure 9: An anomalous behavior of a device detected on January 15^{th}, 2021, due to a drastic reduction in the battery level.

area is flagged with the "broken" status. Figure 9 highlights an abnormal behavior of another device. In this case, the anomalous measurements are due to a drastic reduction in the battery level. At 2 a.m., approximately, the battery died and the device stopped sending data. After changing the battery, at 10 a.m., the device restarts providing reliable measurements.

3.2.3 Gas observations

In the "gas observations" view, a plot for each gas and channel is generated, as shown in Figure 10 for NO. This view meets requirements R3 and R4. The user can choose to visualize the data of the last 24 hours, week or month. The visualization could seem confused, however the user is able to hide one or more lines in the plot thanks to the interactive legend, and zoom in a specific area of the plot. In the web page, the plots related to the two channels of the same gas are placed next to each other (as

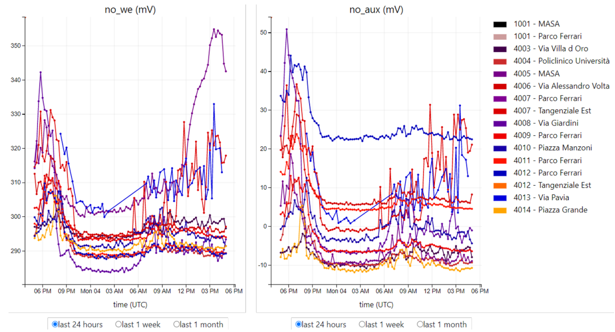

Figure 10: A visualization of the "gas observations" view which shows the measurements of NO channels.

in Figure 10) to facilitate the comparison of these measurements and detect the correlation between the two channels. Thanks to this view, the behavior of the cells can be regularly checked and the maintenance planned.

3.2.4 Sensor anomalies

The accuracy of the raw measurements can be influenced by multiple factors, i.e. the low level of battery voltage, the weather conditions, the air humidity. Distinguishing not correct data allows for providing more reliable data and could improve the results of the calibration task.

We have implemented a majority voting system which combines 3 classifiers: (1) the Sliding Window anomaly detection which considers the consecutive measurements and the IQR to find anomalies far from the normal behavior of the system, (2) the FFIDCAD (Forgetting Factor Iterative Data Capture Anomaly Detection) which is an iterative algorithm, and (3) an algorithm based on the correlation between the values of each gas (NO, NO_2, CO and O3) and the measurements of air temperature and humidity. Every time a new measurement is done by a sensor, just after storing the measurement into the TRAFAIR database, the three classifiers are applied to the measurements.

The research for anomalous data is performed on both channels of each pollutant and device independently, since each device is individual and performs differently from the other devices even if they are in the same location.

The "sensor anomalies" view consists of one plot for each sensor with the raw observations and the anomalies identified by the majority voting system (requirement R6). Also in this case, the user can choose for the observations of the last 24 hours, week, or month.

Figure 11 is an example of anomalies visualization for sensor 4006. Anomalies are identified by a point. As can be seen in the figure, in most cases anomalies are detected in the upper peaks of the time series.

3.2.5 Calibrated observations

The results of the calibration process consists of the concentrations of the 4 measured gases. Starting from 2 values for each gas (one value for each of the two channels) in millivolts, the calibration provides one value in $\mu g/m^3$. Currently, we are using Random Forest to calibrate our data. However, this algorithm can be improved over the time since more and more data are collected and they are used to re-train the calibration algorithm.

The "calibrated observation" view shows the result of the last calibration algorithm, that is the most recent and accurate

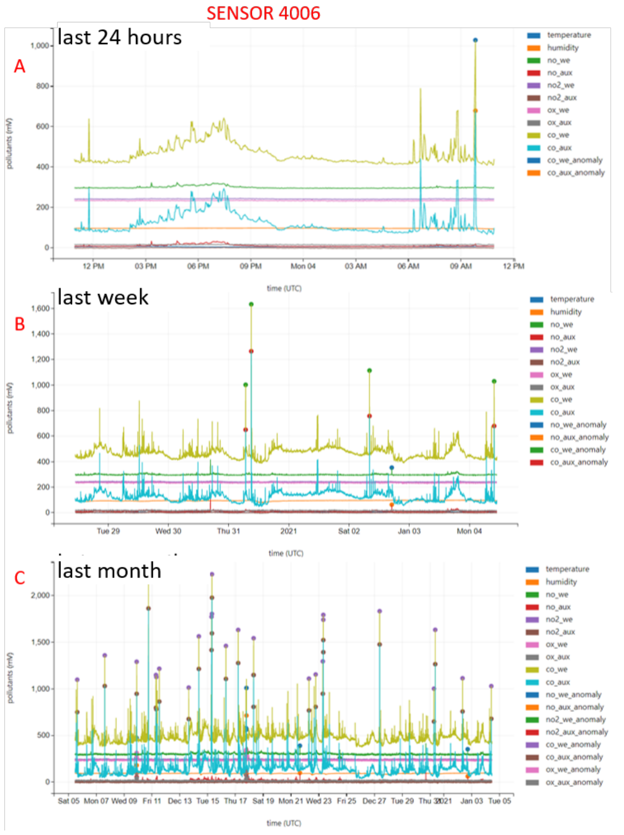

Figure 11: Anomalies of sensor 4006 for the last 24 hours (A), the last week (B), and the last month (C) available on January 4^{th}, 2021 at 11 a.m..

algorithm available for the visualized data. This view meets requirement R5. The calibrated observations are organized in 4 plots, one for each gas, and the user can distinguish the measurements of each device through the integration of the interactive legend.

The calibrated values can be directly compared with the measurements of the AQM stations since they are in the same unit of measure. To validate our calibrated data we have defined one local warning threshold for each gas based on the measurements of the AQM stations. Each threshold has been calculated as $1.25 * M$, where M is the maximum value measured by the AQM stations for the specific gas in the year preceding the date of the observation to be compared. If the concentration of the gas is higher than the corresponding threshold, it is automatically flagged as "anomalous" in the TRAFAIR database by a Python process running in real time. The warning threshold is valid only in the area of Modena since it is provided by the certified values of the AQM stations of Modena and it changes every year. This threshold allows to exclude very high values that are most likely due to malfunction of the sensor. It is not to be confused with the alert thresholds of the European Commission[11] or the reference levels of the European Environment Agency[12], which defines the values to assess the level of pollution in the area.

[11] https://ec.europa.eu/environment/legal/law/5/e_learning/module_2_18.htm
[12] https://www.eea.europa.eu/themes/air/air-quality/resources/air-quality-map-thresholds

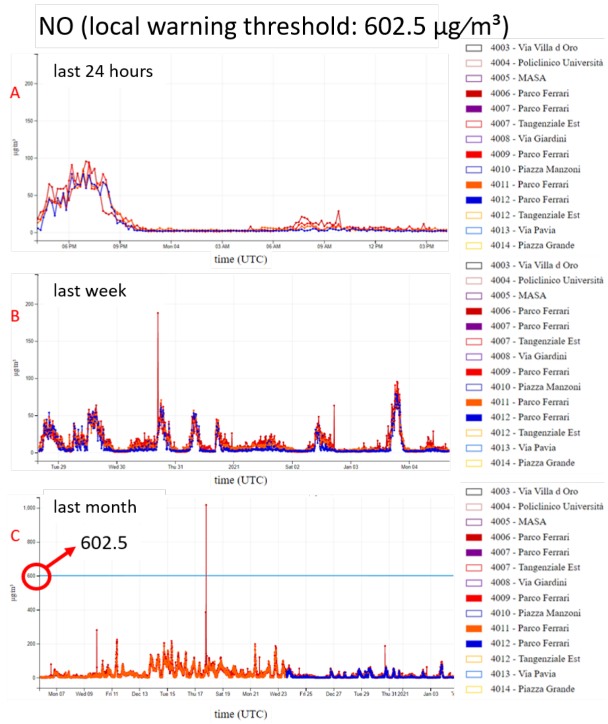

Figure 12: Calibrated *NO* observations by 5 devices located in the same place ("Parco Ferrari") visualized on January 4^{th}, 2021 at 4 p.m..

In the plots of the "calibrated observations" view, a line in correspondence of the threshold value is plotted only if at least one measurement exceeds the threshold. The plots in Figure 12 show the measurements of NO made by 5 different devices installed in the same location named "Parco Ferrari" (this is also the location of an AQM station). We have selected only the devices in the same location through the interactive legend. The concentrations measured by the devices are very similar, as we expected. In the "last month" plot the blue line indicates the above mentioned local warning threshold and only one value is higher than this threshold.

3.2.6 *Certified AQM station measurements*

The sixth view of SenseBoard is devoted to the visualization of AQM station observations. They are hourly certified data related to the concentrations of NO, NO_2, NO_x, and O_3 measured by the two AQM stations installed in Modena (red points in Figure 3).

4 EXPERT EVALUATION

SenseBoard has been regularly used by 4 environmental experts from January 2020 till now and it is still active. It has allowed:

(1) the recording of 250 location/status updates,
(2) the identification of network malfunctions in real time (which occurred twice in the last year and caused the loss of 1-2 days of data),
(3) the detection of sensor faults in semi-real time and anomalous cell behaviour (which occurred 4 times and brought to the cell replacement),
(4) the identification of low battery level which caused anomalous observations (33 times in around 14 months),

(5) the daily comparison of concentrations from low-cost sensors and certified measurements from AQM stations to evaluate the calibration algorithm,
(6) the detection of strange behaviour in the anomaly detection process which allowed to retrain the algorithm and restart it.

The effectiveness of SenseBoard was widely appreciated by environmental engineers who would not have had the opportunity to compare sensor measurements and calibrations and to carry out such sudden checks and maintenance.

5 CONCLUSION

SenseBoard is a data visualization and management platform for air quality sensors. It is a flexible tool that can be integrated into specific IoT environments. In this paper, architecture, users, scope, and exemplar views have been presented. Moreover, details on the sensor data acquisition and storage processes have been given.

SenseBoard is a multi-purpose tool: to manage and maintain the air quality sensor network control and to supervise the calibration process and the identification of anomalies. The management of the network requires the deploy and frequent re-allocation of devices close to the AQM stations or in specific points of interests. Data coming in real-time from the sensors need to be constantly monitored by experts in order to control the normal functioning of sensors.

The dashboard integrates a big amount of heterogeneous data, both geo-spatial and time series data. The position of each sensor is visualized in an interactive map. The measurements of the sensors have been plotted in different line charts with mainly two types of visualization: the same air pollutant measured by all the sensors in the same plot, and all the air pollutants measured by the same sensor in the same plot. Besides, anomalous data are highlighted in other plots. The visualization of such an amount of plots is speed up by the use of Python scripts which generate the plots asynchronously and independently by SenseBoard.

The dashboard is accessible anywhere and anytime to allow a constant monitoring of the network. Besides, it can be generalized to visualize other kinds of geo-spatial and time series data. Indeed, the dashboard is not affected by the type of sensors employed in the network (also in our case we integrate two different types of sensors) and can be easily adapted to monitor other pollutants beyond the ones described in our use case. The flexibility and scalability of SenseBoard allow to monitor networks of a variable number of sensors in cities of different sizes. In addition, in our use case we manage a dynamic sensor network since the sensors are moved frequently. However, this is an additional issue, and the dashboard works also with static sensor networks. SenseBoard can be adapted to query a different data platform which can be a PostgreSQL database or a data model of different type. Queries and plots can be easily modified to visualize data in another way or to show additional data that are not included in our use case.

SenseBoard has been developed according to the technical requirements provided by the environmental experts. Thus, it is not comparable with the dashboards developed for citizens and public administrations. Indeed, the scope of these dashboards is not the monitoring of the sensor network, but the provision of pollution levels to raise awareness among people about the situation in their city. As future work, we will compare SenseBoard with the technical tools provided by the air quality sensor suppliers. In addition, we will integrate an additional view to allow the creation of custom plots, starting from the selection of one or more sensors, pollutants, and AQM stations, and the time interval. This will allow for further data comparison.

ACKNOWLEDGMENTS

Research reported in this paper was partially supported by the TRAFAIR project 2017-EU-IA-0167, co-financed by the Connecting Europe Facility of the European Union. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of EU Commission. The authors would like to thank the City of Modena that contributes to the deployment of the LoRa network, and the LARMA research group for providing requirements and feedback on SenseBoard. Moreover, a special thank goes to ARPAE that shared real-time air quality observations used for the calibration of the low-cost air quality sensors.

REFERENCES

[1] European Environment Agency. 2020. Air quality in Europe — 2020 report. Issue 9. https://doi.org/10.2800/786656 Available at https://www.eea.europa.eu//publications/air-quality-in-europe-2020-report.
[2] United Nations General Assembly. 2015. Transforming our world: The 2030 Agenda for Sustainable Development. Available at http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.
[3] Chiara Bachechi, Federica Rollo, Federico Desimoni, and Laura Po. 2020. Using Real Sensors Data to Calibrate a Traffic Model for the City of Modena. In *Intelligent Human Systems Integration 2020*, Tareq Ahram, Waldemar Karwowski, Alberto Vergnano, Francesco Leali, and Redha Taiar (Eds.). Springer International Publishing, Cham, 468–473.
[4] Titus Balan, Catalin Dumitru, Gabriela Dudnik, Enrico Alessi, Suzanne Lesecq, Marc Correvon, Fabio Passaniti, and Antonella Licciardello. 2020. Smart Multi-Sensor Platform for Analytics and Social Decision Support in Agriculture. *Sensors* 20, 15 (2020), 4127. https://doi.org/10.3390/s20154127
[5] A. Bigi, G. Veratti, S. Fabbi, L. Po, and G. Ghermandi. 2019. Forecast of the impact by local emissions at an urban micro scale by the combination of Lagrangian modelling and low cost sensing technology: The TRAFAIR project. *19th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Harmo 2019* (2019). https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084160462&partnerID=40&md5=3b37303a7af769206777d87ab30a2541 cited By 2.
[6] Mehmet Ali Ertürk, Muhammed Ali Aydin, M. Talha Buyukakkaslar, and Hayrettin Evirgen. 2019. A Survey on LoRaWAN Architecture, Protocol and Technologies. *Future Internet* 11, 10 (2019), 216. https://doi.org/10.3390/fi11100216
[7] N. Kumar, D. Acharya, and D. Lohani. 2021. An IoT-Based Vehicle Accident Detection and Classification System Using Sensor Fusion. *IEEE Internet of Things Journal* 8, 2 (2021), 869–880. https://doi.org/10.1109/JIOT.2020.3008896
[8] Xiangtao Liu, Tianle Zhang, Ning Hu, Peng Zhang, and Yu Zhang. 2020. The method of Internet of Things access and network communication based on MQTT. *Computer Communications* 153 (2020), 169 – 176. https://doi.org/10.1016/j.comcom.2020.01.044
[9] Luis F. Luque-Vega, David A. Michel-Torres, Emmanuel López-Neri, Miriam A. Carlos-Mancilla, and Luis Enrique González Jiménez. 2020. IoT Smart Parking System Based on the Visual-Aided Smart Vehicle Presence Sensor: SPIN-V. *Sensors* 20, 5 (2020), 1476. https://doi.org/10.3390/s20051476
[10] L. Po, F. Rollo, C. Bachechi, and A. Corni. 2019. From Sensors Data to Urban Traffic Flow Analysis. In *2019 IEEE International Smart Cities Conference (ISC2)*. IEEE, Casablanca, Morocco, 478–485. https://doi.org/10.1109/ISC246665.2019.9071639
[11] L. Po, F. Rollo, J. R. R. Viqueira, R. T. Lado, A. Bigi, J. C. López, M. Paolucci, and P. Nesi. 2019. TRAFAIR: Understanding Traffic Flow to Improve Air Quality. In *2019 IEEE International Smart Cities Conference (ISC2)*. IEEE, Casablanca, Morocco, 36–43. https://doi.org/10.1109/ISC246665.2019.9071661
[12] N. Shivaraman, S. Saki, Z. Liu, S. Ramanathan, A. Easwaran, and S. Steinhorst. 2020. Real-Time Energy Monitoring in IoT-enabled Mobile Devices. In *2020 Design, Automation Test in Europe Conference Exhibition (DATE)*. IEEE, Grenoble, France, 991–994. https://doi.org/10.23919/DATE48585.2020.9116577
[13] D. Zhang and S. S. Woo. 2020. Real Time Localized Air Quality Monitoring and Prediction Through Mobile and Fixed IoT Sensing Network. *IEEE Access* 8 (2020), 89584–89594. https://doi.org/10.1109/ACCESS.2020.2993547