Air Quality Sensor Network Data Acquisition, Cleaning, Visualization, and Analytics: A Real-world IoT Use Case

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

Bharath Sudharsan
Data Science Institute,
NUIG, Galway, Ireland
b.sudharsan1@nuigalway.ie

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

John Breslin
Data Science Institute,
NUIG, Galway, Ireland
john.breslin@nuigalway.ie

2021UBICOMP
September 21 — 26
All Over the World

Introduction

Urban air pollution is a serious public health hazard in many EU cities. Monitoring and analyzing air quality is of primary importance to encourage more sustainable lifestyles and plan corrective actions. The scope of this poster is to describe the design and end-to-end implementation of a real-world urban air quality data collection and analytics use case started in August 2019 in Modena city, Italy. The work is a part of the **TRAFAIR** European project (www.trafair.eu).

Implementation

Goal: Hyper-local monitoring air quality in multiple location points

Infrastructure:

- 13 low-cost IoT devices installed in 12 locations in Modena
- 4 cells (sensors), one for each gas (**NO**, **NO2**, **CO** and **Ox**)

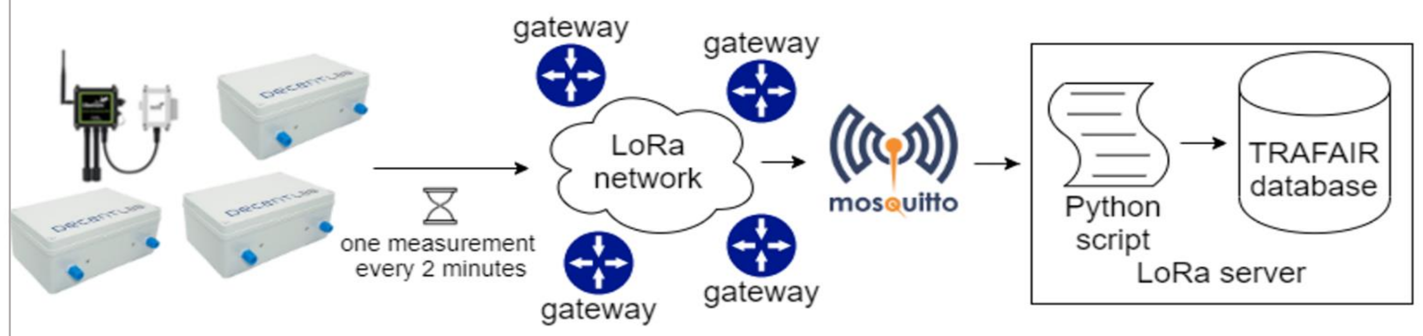
Data:

- 1 measure for each gas and channel (we and aux) in *mV*
- 1 measure for temperature and humidity

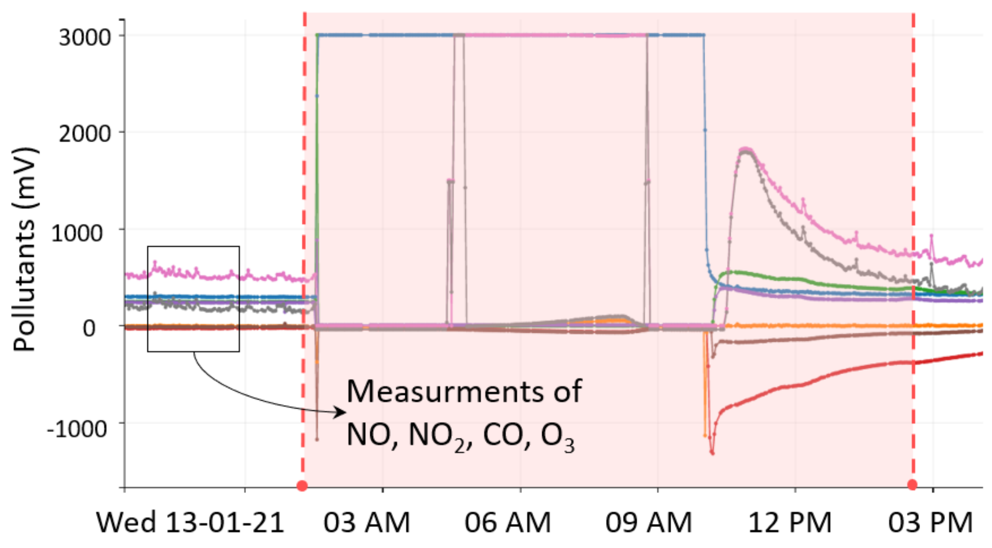
Challenge: Low-cost devices are less reliable than the Air Quality Monitoring legal stations and require a complex calibration process to obtain pollutant concentration data from the raw observations in *mV*.

Data Acquisition, Data Cleaning and Calibration

IoT devices are connected to the LoRaWan network through gateways and transmit data every 2 minutes. When the gateways receive data, they send them to the LoRa server where the MQTT (Message Queue Telemetry Transport) Broker Mosquitto is running. Then, the data are parsed in a python script and stored in a PostgreSQL database.



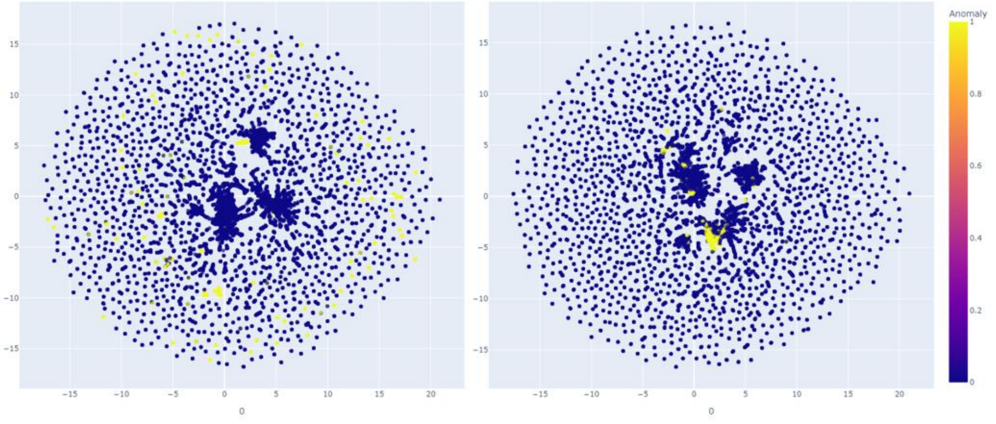
Anomaly detection algorithms are executed to identify and remove such abnormal data patterns (below Figure). Repo with code released at: https://github.com/bharathsudharsan/Air-Quality-IoT-Analytics



The output of anomaly detection is clean data that are employed by the calibration process to generate reliable pollutant concentration. Calibration exploits the correlation between the raw measurements from low-cost devices and the accurate values from the legal stations.

Unsupervised Anomaly Detection Models

Models created using PyCaret, trained using part of TRAFAIR dataset



Umap for outliers: yellow points in plot are outliers

Remove anomalies: Run any provided model. Then, remove the data rows that correspond to high anomaly scores (set a threshold).

Conclusion

The deployed implementation allows building a **historical air quality dataset** with **accurate** (calibrated with great precision close to AQM stations) and **reliable** (resilient LoRa networks) concentrations of air pollutants.

Raw and calibrated data are available as **open data** on the open data portal of the Emilia Romagna region https://dati.emilia-romagna.it/ and the National and European data portals.

References

L. Po, F. Rollo, J. R. R. Viqueira, et al.2019. TRAFAIR: Understanding Traffic Flow to Improve Air Quality. In IEEE International Smart Cities Conference (ISC2).

Federica Rollo and Laura Po. 2021. SenseBoard: Sensor Monitoring for Air Quality Experts. In 24th International Conference on Database Theory.

Bharath Sudharsan, John G Breslin, and other. 2021. Porting and Execution of Anomalies Detection Models on Embedded Systems in IoT: Demo Abstract. In International Conference on Internet-of-Things Design and Implementation (IoTDI).