On Assessing the Accuracy of Air Pollution Models Exploiting a Strategic Sensors Deployment

Rita Tse
Macao Polytechnic Institute, China
ritatse@ipm.edu.mo

Lorenzo Monti
University of Bologna, Italy
lorenzo.monti20@unibo.it

Catia Prandi
University of Bologna, Italy
catia.prandi2@unibo.it

Davide Aguiari
Università degli studi di Cagliari, Italy
davide.aguiari@unica.it

Giovanni Pau
University of Bologna, Italy
giovanni.pau@unibo.it

Paola Salomoni
University of Bologna, Italy
paola.salomoni@unibo.it

ABSTRACT

This paper presents a preliminary experiment done to identify potential problems and issues in setting up a testbed for air pollution measurement and modeling. Our final testbed, part of a joint research activity between the University of Bologna and the Macao Polytechnic Institute, will be composed of three lines of the air pollution sensors Canarin II and it will be used to produce spatio-temporal open data to test third-party air pollution models. Here, we present a preliminary experiment based on a single line of sensors, showing interesting insights into the actual open challenge of air pollution modeling techniques validation, taking into account the effects of air pollutant emissions sources, meteorology, atmospheric concentrations and urban vegetation.

CCS CONCEPTS

• **Applied computing** → Environmental sciences;

KEYWORDS

Air quality sensors, air pollution modeling, sensors deployment, open data

ACM Reference Format:
Rita Tse, Lorenzo Monti, Catia Prandi, Davide Aguiari, Giovanni Pau, and Paola Salomoni. 2018. On Assessing the Accuracy of Air Pollution Models Exploiting a Strategic Sensors Deployment. In *International Conference on Smart Objects and Technologies for Social Good (Goodtechs '18), November 28–30, 2018, Bologna, Italy.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3284869.3284880

1 INTRODUCTION

Air pollution is a phenomenon by which solid and liquid particles and gases contaminate the environment with negative effects on population health [6, 14]. The effects can be really serious, even lethal, as stated by the World Health Organization (WHO) that renominated this phenomenon *the invisible killer*, estimating in 7 million the number of deaths every year caused by the exposure to fine particles in polluted air [13]. In the same report WHO also claimed that 9 out of 10 people worldwide breathe polluted air and more than 80% of people live in urban areas where the air quality levels exceed the WHO guideline level [13]. Different strategies can be employed to tackle this problem and achieving sustainable development, such as sustainable transport, more efficient and renewable energy production and use and waste management [12]. The first action that local governments and policymakers should tackle is the gathering of air quality data to reflect their commitment to air pollution assessment and monitoring [21]. Having the data, it becomes relevant to develop models able to understand and predict the way pollutants behave in the atmosphere, so as to state the actual air quality in an area [20], and then equipping citizens with tailored services (i.e., mobile apps computing personalized pedestrian and cycling paths in the urban environment [9, 15], keeping the citizens in mind, as well as their preferences and needs [8, 10]).

Several mathematics theories and numerical tools have been studied in the literature, under the umbrella term *air pollution modeling*, to understand the causal relationship between emissions, meteorology, atmospheric concentrations, deposition, and other factors [18]. The techniques used are several (see, for example, [3, 7]) but the common goal is to make an assessment of pollutant impact over a given area using a defined set of data.

In the last years, new technologies are emerging, which provide an alternative and cost-efficient approach to measure air pollutants [17]. Being low-cost and adequately precise, this new generation of pollution sensors are completely changing the possibility to be aware of the air quality in the urban environment. In this context, we are involved in the development and test of Canarin II, the result from the collaboration among the University of Bologna, the Universitè Pierre et Marie Curie, the Macao Polytechnic Institute, and the Asian Institute of Technology. The Canarin II architecture works on a UDOO Neo Full, an Arduino-powered Android/Linux single board. It measures PM1, PM2.5 and PM10, pressure, temperature, humidity and UV [1].

The availability of Canarin II sensors drove us in designing and setting up an air quality testbed, with three goals in mind:

(i) To produce a set of spatio-temporal open data in different weather condition in order to provide third parties with air quality data to test their own models. Other measures (e.g. sensed wind, detailed map of the area, vegetation distribution) will be provided in order to offer a complete set of data to appropriately test pollution diffusion models.

Goodtechs '18, November 28–30, 2018, Bologna, Italy

ACM ISBN 978-1-4503-6581-9/18/11.
https://doi.org/10.1145/3284869.3284880

(ii) To develop models to measure and test the accuracy and efficacy of air pollution modeling techniques, using outdoor air quality data collected under the presence of specific circumstances that can affect the outcome, such as present and future barriers. To make the defined models even stronger, in the future, such dataset could be integrated with data recorded using mobile sensors provided to users and gathered while moving in the surrounding area (e.g. using sensors on shared bikes [1]).
(iii) To determine the best configuration needed in term of the number of sensors and distance between each sensor in order to collect air quality data and assess with a high accuracy the validity of air pollution models.

To address the three research issues, we designed a preliminary experiment using a set of sensors that are being deployed around a new building, 30.000 m2 wide, which is the new seat of the Campus of Cesena of the University of Bologna (in the Cesena city). The building is located in a particular area, surrounded both by pollutant sources (e.g. a highway and a railway) and by residential/green areas. Due to the peculiar characteristics of the area and the shape of the building, a strategic sensors deployment will let us collect a rich dataset that incorporates phenomena affecting the air quality assessment.

2 AIR POLLUTION MODELS ASSESSMENT

As briefly mentioned, several air pollution modeling techniques have been studied and developed to address three main concerns: to assess the existing air quality situation and calculate the population exposure to pollution; to forecast changes in pollution levels and prevent or inform about oncoming predicted critical episodes; to define an air quality planning program. Often these models have been used i) in various forms, ii) with not-standardized and not accurate enough datasets, and ii) with differing and often incomparable quality assurance methods, at both national and local levels. This scenario let emerge the urgent need to harmonize the way these models are validated so as to achieve reliable results.

Different factors affect the outcome resulting from the application of air pollution modeling techniques and need to be considered in the design of the models to avoid uncertain results [5]. Such issues include:

- urban vegetation: vegetation can, directly and indirectly, affect local and regional air quality by altering the urban atmospheric environment [11];
- background concentration: this factor indicates the concentration that would be measured if local sources were not present [2];
- urban layout: buildings can alter the concentration and deposition values [3];
- terrain: the conformation of the area needs to be considerate to simulate the movement of pollutants in the atmosphere [3];
- water source: water, in form of rivers, lakes, or oceans, may transport pollution for long distance, and, sometimes, in high concentrations [4];

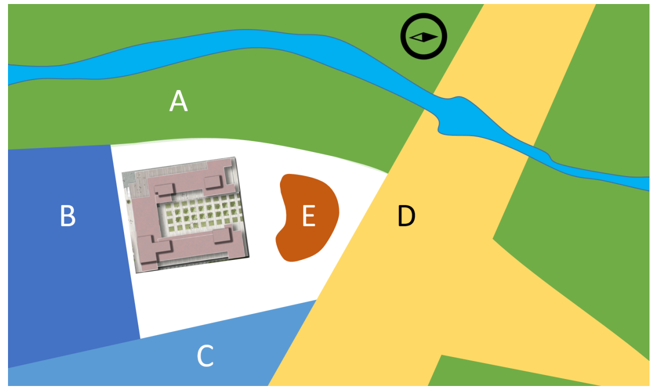

Figure 1: Areas around the campus building.

- meteorological data: information about wind speed and direction, temperature, humidity are relevant in air pollution modeling [16].

Acquiring precise and variegated enough input data to test the models is a hard task. Some projects have been created with the aim of creating an open data collection of air quality data. One example is represented by OpenAQ[1] with a dataset aggregating air quality measurements, obtained by government agencies, from 8,589 locations in 67 countries. Even than this project is really interesting and has hight potential as a tool to inform people about monitored pollution, the dataset can not be used for air pollution modeling since it doesn't include important information related to the context the data are collected, neither the accuracy of the used sensors. Unfortunately, this problem is common to several government open data air quality measures collections. For this reason, with this experiment we want to fill the gap providing an open dataset of spatio-temporal data, providing all the information needed to assess in a rigorous way the accuracy of thirty-party air pollution models.

3 THE SPATIAL CONTEXT

Our testbed is being set up in the Campus of Cesena of the University of Bologna. It is a new building (partially under construction), located at the border between a park (see A in Figure 1, west direction), a residential area (B, sud), a not operating industrial zone (C, east) and the railroad/highway area (D, north). The park is part of the Savio River Reserve, which extends along the River Savio in the part where the reserve enters the urban area of Cesena[2]. The not-operating industrial area is occupied by two buildings originally devoted to fruit storage and distribution. The highway and the railroad enter the city of Cesena going parallel. The highway goes underground to cross the city center nearby the campus building (300 m). The railroad reaches the station 1,2 Km after the campus building. In order to protect the building from pollution coming from the railroad and the highway, an artificial hill (E, north) was created in the north garden, between the building itself and the railroad/highway area.

[1]https://openaq.org/
[2]https://en.wikipedia.org/wiki/Savio_River_Reserve

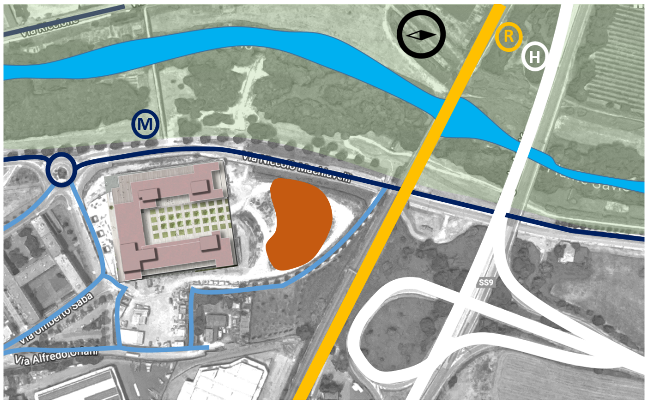

Figure 2: Detail of roads around the building.

Figure 3: The artificial hill created to protect the building from pollution coming from the railroad and the highway.

Figure 2 shows a more detailed map, where the main sources of pollution are depicted respectively in white (H, the highway) and yellow (R, the railroad). A minor source of pollution is via Macchiavelli (M, in blue in figure 1). All other streets around the building (depicted in light blue) can be considered irrelevant due to the very limited traffic flow.

The artificial hill E partially protects the building from pollution coming from the D area. Figure 3 shows the view of the railroad and the highway taken from the internal part of the courtyard garden. While the hill protects the right part of the view, on the left both the railway and the highway are visible. Similar pictures can be taken from the east side of the building.

4 THE SENSORS STATION

The Canarin II [1] is the result of the collaboration among the Macao Polytechnic Institute, the Asian Institute of Technology, and the Pierre and Marie Curie Sorbonne University.

The sensors station senses different air quality and environmental conditions, such as PM 1.0 (<10 microm) particles, PM 2.5 (<2.5 microm) particles, and PM 10 (<10 microm) particles, temperature, relative humidity, and air pressure.

The architecture is based on UDOO Neo Full [3], an Arduino-powered Android/Linux single board computer that overcomes some limitations emerged in the first version of the sensors station [19], such as:

- a new 1GHz ARM Cortex-A9 microprocessor;
- the inclusion of PM 1.0 (<1.0 microm) particles concentration and formaldehyde sensors;
- a more powerful SD card;
- a new Wi-Fi module for a stronger connection stability;
- a better management of the sensed data: if the Internet connection is missing, the data are saved on the local SD and send to the server when possible.

The new system architecture is, therefore, structured around three layers: (i) all the sensors run on an Arduino UNO-compatible platform that clocks at 200 MHz, based on a Cortex-M4 I/O real-time co-processor; (ii) a Linux based OS stores the data into files and it establishes a connection to the server in order to send data; (iii) a server side that provides a data repository and data intelligence as well. The sensors station design is based on an in-house printed circuit board (PCB) that hosts the board and the sensors welded. The communication is based on Wi-Fi and the board is also configured to be connected to GSM network using the EAP-SIM (EAP Subscriber Identity Module) authentication framework by means of a USB SIM reader. The communication protocol is based on a customized UDP and it communicates to an enhanced MySQL database version.

We also designed an in-house 3D printed PLA (polylactic acid) box to wrap the battery, the PCB, and the sensors together. Everything fits in a 19x15x7 cm and it weighs about 900g due to the battery and the enclosure; a smaller and more portable 3D printed box is about to be produced.

5 THE SENSORS DEPLOYMENT

As shown in Figure 4, our testbed will be set up using three outdoor spaces: the west terrace, facing via Macchiavelli (1.WT - located at second floor of the campus building), the courtyard garden (2.CG - located at the ground floor) and the east terrace (3.ET - located at second floor of the campus building). Canarin II pollution sensors station will be located in three parallel lines of 4, one every 20 meters. The testbed is completed by an anemometer, to measure the speed of the wind, located on the rooftop of the building (4.AN, located at the roof of the third floor of the campus building).

The setup of the whole testbed will be completed in the next months, accordingly with the completion of the building construction. This paper presents some preliminary tests conducted on a partial sensors' infrastructure, located in the WT (this area of the building has been already completed). The dislocation of one line of sensors stations in the WT area is shown in Figure 5, with a distance of 20 meters between each Canarin II. Having four sensors station in line allow as to overcome issues related to the accuracy and calibration of the specific sensor. We estimated in 20 meters the right distance to collect data affected by different factors. In the next future, more tests will be made to define the best distance needed between each sensor, and if this value is affected by the position of the sensors in the different areas (i.e., WT, GT and ET), settled to capture different urban scenarios and pollution sources.

In a month of data gathering, we already gathered 714,240 complex entries in our database. Each entry includes several values (such as PM 1.0, PM 2.5, PM 10, humidity, temperature, wind direction and speed, air pressure) contextualized in space (the GPS coordinates) and time.

[3] https://www.udoo.org/udoo-neo/

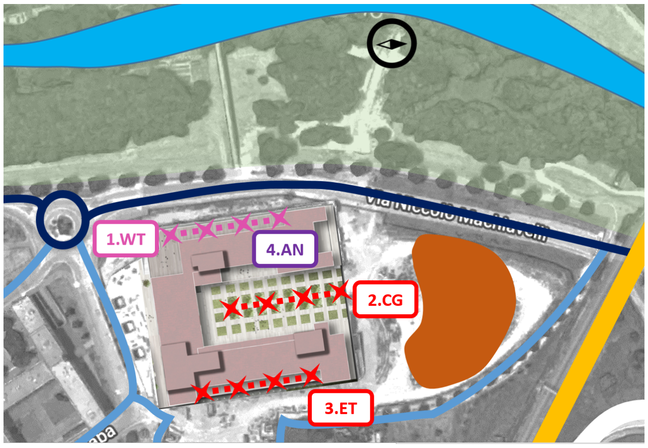

Figure 4: Testbed design and experimental setup.

Figure 5: The deployment of one line of sensors

6 CONCLUSION AND FUTURE WORKS

In this paper, we present a preliminary experiment to strategically deploy sensors stations in a new University campus. At this stage, the campus is composed by a single building that is located in a peculiar area, allowing to sense air quality data affected by different factors, including vegetation, different mobile pollution sources, a specific urban layout, and the proximity to a river natural reserve.

With this research project, we intend to address three open issues: i) producing an accurate open data collection that can be exploited by third parties to assess their own air pollution models; ii) developing models to measure the accuracy and test the efficiency of air pollution modeling techniques; iii) defining strategic deployment configurations of air quality and environmental sensors (e.g., the distance between each sensor, the number of sensors, positions).

As an initial deployed, we used a line composed by four Canarin II sensors stations. Each sensors station is able to accurately sense PM 1.0, PM 2.5, PM 10, relative humidity, temperature. Moreover, we augmented the sensing platform with an anemometer, to measure the speed of the wind. In a month, we already collected 714,240 spatio-temporal entries, where each entry is composed of all the sensed environmental conditions at a given time and location.

We are planning to extend the actual deployment including other two lines of sensors located in strategic positions in the building, facing areas with different characteristic (such as vegetation, layout, exposure to pollutant sources). Moreover, we are considering to integrate our dataset with data gathered using mobile air quality sensors to create an even more accurate and variegated collection of open data.

ACKNOWLEDGMENTS

This work was supported in part by the Macao Polytechnic Institute - Indoor Environmental Monitoring System for Heritage Building at Coimbra (RP/ESAP-03/2017).

REFERENCES

[1] Davide Aguiari, Giovanni Delnevo, Lorenzo Monti, Vittorio Ghini, Silvia Mirri, Paola Salomoni, Giovanni Pau, Marcus Im, Rita Tse, Mongkol Ekpanyapong, et al. 2018. Canarin II: Designing a smart e-bike eco-system. In *Consumer Communications & Networking Conference (CCNC), 2018 15th IEEE Annual*. IEEE, 1–6.
[2] Saravanan Arunachalam, Alejandro Valencia, Yasuyuki Akita, Marc L Serre, Mohammad Omary, Valerie Garcia, and Vlad Isakov. 2014. A method for estimating urban background concentrations in support of hybrid air pollution modeling for environmental health studies. *International journal of environmental research and public health* 11, 10 (2014), 10518–10536.
[3] Aaron Daly and Paolo Zannetti. 2007. Air pollution modeling–An overview. *Ambient air pollution* (2007), 15–28.
[4] E Wendell Hewson and Lars E Olsson. 1967. Lake effects on air pollution dispersion. *Journal of the Air Pollution Control Association* 17, 11 (1967), 757–761.
[5] Piotr Holnicki and Zbigniew Nahorski. 2015. Emission data uncertainty in urban air quality modeling—case study. *Environmental Modeling & Assessment* 20, 6 (2015), 583–597.
[6] Marilena Kampa and Elias Castanas. 2008. Human health effects of air pollution. *Environmental pollution* 151, 2 (2008), 362–367.
[7] Rong Lu, Richard P Turco, and Mark Z Jacobson. 1997. An integrated air pollution modeling system for urban and regional scales: 1. Structure and performance. *Journal of Geophysical Research: Atmospheres* 102, D5 (1997), 6063–6079.
[8] Silvia Mirri, Catia Prandi, Marco Roccetti, and Paola Salomoni. 2017. Handmade narrations: handling digital narrations on food and gastronomic culture. *Journal on Computing and Cultural Heritage (JOCCH)* 10, 4 (2017), 20.
[9] Silvia Mirri, Catia Prandi, Marco Roccetti, and Paola Salomoni. 2017. Walking under a Different Sky: Urban Colored Routes for Creative Engagement and Pleasure. *International Journal of Human–Computer Interaction* 33, 12 (2017), 1010–1021.
[10] Silvia Mirri, Marco Roccetti, and Paola Salomoni. 2018. Collaborative design of software applications: the role of users. *Human-centric Computing and Information Sciences* 8, 1 (2018), 6.
[11] David J Nowak, Patrick J McHale, Myriam Ibarra, Daniel Crane, Jack C Stevens, and Chris J Luley. 1998. Modeling the effects of urban vegetation on air pollution. In *Air pollution modeling and its application XII*. Springer, 399–407.
[12] Abdeen Mustafa Omer. 2008. Energy, environment and sustainable development. *Renewable and sustainable energy reviews* 12, 9 (2008), 2265–2300.
[13] World Heath Organization. 2018. 9 out of 10 people worldwide breathe polluted air, but more countries are taking action. http://www.who.int/news-room/detail/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-coun\tries-are-taking-action. [Online; accessed 19-June-2018].
[14] C Arden Pope, Michael J Thun, Mohan M Namboodiri, Douglas W Dockery, John S Evans, Frank E Speizer, Clark W Heath, et al. 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. *American journal of respiratory and critical care medicine* 151, 3 (1995), 669–674.
[15] Catia Prandi, Andrea Melis, Marco Prandini, Giovanni Delnevo, Lorenzo Monti, Silvia Mirri, and Paola Salomoni. 2018. Gamifying cultural experiences across the urban environment. *Multimedia Tools and Applications* (2018), 1–24.
[16] Nelson L Seaman. 2000. Meteorological modeling for air-quality assessments. *Atmospheric environment* 34, 12-14 (2000), 2231–2259.
[17] Emily G Snyder, Timothy H Watkins, Paul A Solomon, Eben D Thoma, Ronald W Williams, Gayle SW Hagler, David Shelow, David A Hindin, Vasu J Kilaru, and Peter W Preuss. 2013. The changing paradigm of air pollution monitoring.
[18] Anjali Srivastava and B Padma S Rao. 2011. Urban air pollution modeling. In *Air Quality-Models and Applications*. InTech.
[19] Rita Tse and Giovanni Pau. 2016. Enabling street-level pollution and exposure measures: a human-centric approach. In *Proceedings of the 6th ACM International Workshop on Pervasive Wireless Healthcare*. ACM, 1–4.
[20] PEAN UNION et al. 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. *Official Journal of the European Union* (2008).
[21] Daniel A Vallero. 2016. Air pollution monitoring changes to accompany the transition from a control to a systems focus. *Sustainability* 8, 12 (2016), 1216.