

Research Article

An Architecture Offering Mobile Pollution Sensing with High Spatial Resolution

Oscar Alvear, Willian Zamora, Carlos Calafate, Juan-Carlos Cano, and Pietro Manzoni

Department of Computer Engineering (DISCA), Universitat Politècnica de València, 46022 Valencia, Spain

Correspondence should be addressed to Oscar Alvear; osal@doctor.upv.es

Received 21 December 2015; Revised 20 February 2016; Accepted 14 March 2016

Academic Editor: Liling Fu

Copyright © 2016 Oscar Alvear et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mobile sensing is becoming the best option to monitor our environment due to its ease of use, high flexibility, and low price. In this paper, we present a mobile sensing architecture able to monitor different pollutants using low-end sensors. Although the proposed solution can be deployed everywhere, it becomes especially meaningful in crowded cities where pollution values are often high, being of great concern to both population and authorities. Our architecture is composed of three different modules: a mobile sensor for monitoring environment pollutants, an Android-based device for transferring the gathered data to a central server, and a central processing server for analyzing the pollution distribution. Moreover, we analyze different issues related to the monitoring process: (i) filtering captured data to reduce the variability of consecutive measurements; (ii) converting the sensor output to actual pollution levels; (iii) reducing the temporal variations produced by mobile sensing process; and (iv) applying interpolation techniques for creating detailed pollution maps. In addition, we study the best strategy to use mobile sensors by first determining the influence of sensor orientation on the captured values and then analyzing the influence of time and space sampling in the interpolation process.

1. Introduction

Air pollution basically consists in the emission of gases or particles into the atmosphere, producing changes in its composition. Air pollution levels are critical aspect to consider nowadays since it is associated with several problems affecting people's life quality, such as health issues (mainly in the respiratory tracts), climate changes, and reduced agriculture production.

Throughout Europe, about one thousand five hundred air monitoring stations have been deployed to control air pollution on a large scale, providing coarse-granularity pollution levels for most relevant cities. Despite the fact that this number may seem large, when focusing on a specific city we find that these stations are quite scarce, failing to provide detailed pollution levels on a per-neighborhood basis [1]. For instance, in Valencia, the third largest city in Spain, there are only 5 monitoring stations, as shown Figure 1. These are able to provide pollution levels on a large scale but for better data granularity and to study spatial variability with more detail, it would be necessary to have many of these stations, which becomes unfeasible due to the high costs associated.

Monitoring stations rely on sophisticated sensors, which are very accurate and introduce minimum uncertainty levels in the data capture process (e.g., Dobson spectrophotometers are used for monitoring ozone levels [2]). However, they are very expensive and hard to manage. Due to their size, they must be installed on a specific location, and the monitored value is only representative in a small surrounding area.

An alternative for measuring environmental pollution is relying on mobile sensing. Specifically, small low-cost devices can be installed in various types of vehicles to monitor different parts of the city at different times. The main problem of low-end mobile sensors is that they have less accuracy than sophisticated sensors, and so they need to be regularly calibrated; besides, measurements are also weather-dependent.

Even a small and low-cost mobile station must be endowed with several sensors able to measure different types of air pollution. Pollutants can be of two types: (i) primary air pollutants, which are gases or particles emitted directly into the atmosphere: in this category we have carbon monoxide (CO), carbon dioxide (CO₂), particulate matter smaller than 10 microns (PM10), or particulate matter smaller than 2.5

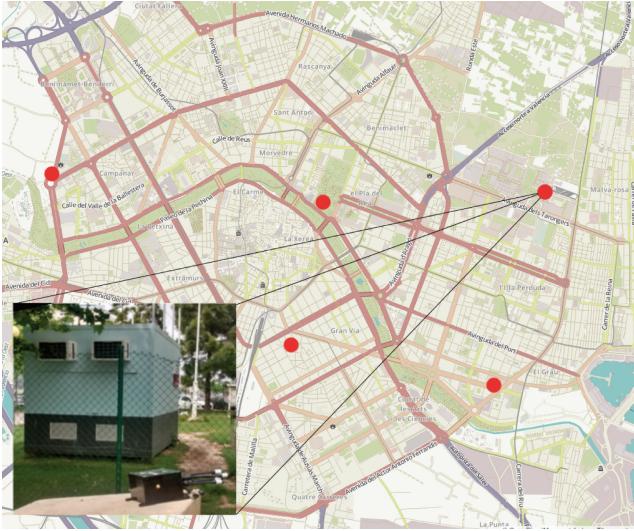


FIGURE 1: Example of an air monitoring station, and the location of the 5 stations available in Valencia, Spain.

microns (PM_{2.5}); and (ii) secondary air pollutants, which are gases produced by a chemical reaction between primary pollutants and some environment element: in this second category we have ozone (O₃), which is produced by the combination of nitrogen oxides (NO_x), Oxygen (O₂), Volatile Organic Compounds (VOC), and sunlight [3].

In this paper we propose an architecture offering mobile pollution sensing with high spatial resolution. Our architecture includes three independent modules: a mobile sensor for monitoring environment pollutants, an Android-based device for transferring the gathered data to a central server, and a central processing server for analyzing the pollution distribution using the collected data through spatial interpolation techniques. Throughout the paper, we will focus on ozone sensing since it is more complex to estimate than other pollutants. In particular, we will discuss how to properly calibrate the sensor and reduce time variability. In addition, we will assess the impact of sensor position and of mobility, as well as the impact of temporal and spatial subsampling.

The paper is organized as follows. In the next section we present some related works on the topic. In Section 3, we provide an overview of the proposed architecture. Then, in Section 4, we detail the procedure followed in order to obtain reliable measurements from low-cost sensing devices. Section 5 discusses the optimal strategy for performing mobile measurements. In Section 6, we validate the proposed architecture through comparison against infrastructure-based results. Finally, in Section 7, we present our conclusions and future work.

2. Related Work

In recent decades, air pollution monitoring has gained worldwide relevance due to the influence of air quality on our lives. There are many research works that study the effects of pollution on our health. Among them we can find the contributions of Chen et al. [4, 5], who analyzed the effects

of ozone and particles matter on human health. Brook et al. [6] also contributed to this field by studying the relationship between the exposure to air pollution (including ozone) and cardiovascular events.

Determining the pollution distribution in a city based on a few samples requires adopting spatial interpolation techniques for estimating it. In this regard, studies such as [7, 8] have relied on kriging interpolation techniques to predict pollution. These studies were made in the cities of Quebec and Toronto, respectively.

To have a detailed overview of pollution distribution, fine-grain monitoring is required, and mobile sensing is the best option to achieve it. In the literature we can find several works adopting this approach. For instance, Brković and Sretović [9] propose a system to monitor environment pollution in the city of Belgrade using Waspmote sensors installed in the public transport system. Deng and Zhang [10] use a vehicular sensor network for air pollution monitoring. In particular, they propose to use taxis for deploying the system and mainly analyze the communication between them. More recently, Calafate and Duroth [11] combined mobile sampling techniques with kriging-based interpolation to determine the achievable accuracy when estimating the ozone distribution in a city, relying on the public transportation system for data gathering.

Cheng et al. [12] propose a system to monitor the concentrations of PM_{2.5} using crowdsourcing, which is an alternative to using mobile sensors. They focus on the analysis of the mechanical sensor design to optimize the air reception, as well as on data fusion techniques to analyze the data. Sensor calibration is achieved by analyzing data produced in the laboratory using neural networks.

Finally, Zheng et al. [13] show how to analyze the data obtained from different sources, such as traffic levels, weather conditions, and pollution, using different Big Data techniques, and evidence how these techniques allow inferring environmental pollution levels with better granularity.

Our proposal differs from the former ones since it aims at providing a full mobile sensing architecture. In particular, we combine low-end sensors, smartphones, and Cloud services to efficiently monitor pollution levels. By relying on the data readings provided by the existent infrastructure (high reliability), we show how to calibrate and adjust data readings and how it is possible to obtain detailed pollution maps using spatial interpolation techniques. In addition, we study the impact of mobility, sensor orientation, spatial subsampling, and time subsampling on the prediction accuracy.

3. Mobile Sensing Architecture Overview

Our proposed architecture defines a set of elements that allow monitoring air pollution in a cheap and easy way, being specially useful in very crowded cities. It combines information from existing air quality monitoring stations with the data collected by mobile sensors to generate fine-grained reports about spatial pollution distribution throughout the city. These mobile sensors can be installed in bicycles or the public transportation system to monitor the whole city in a simple and effective way. All collected information is stored on

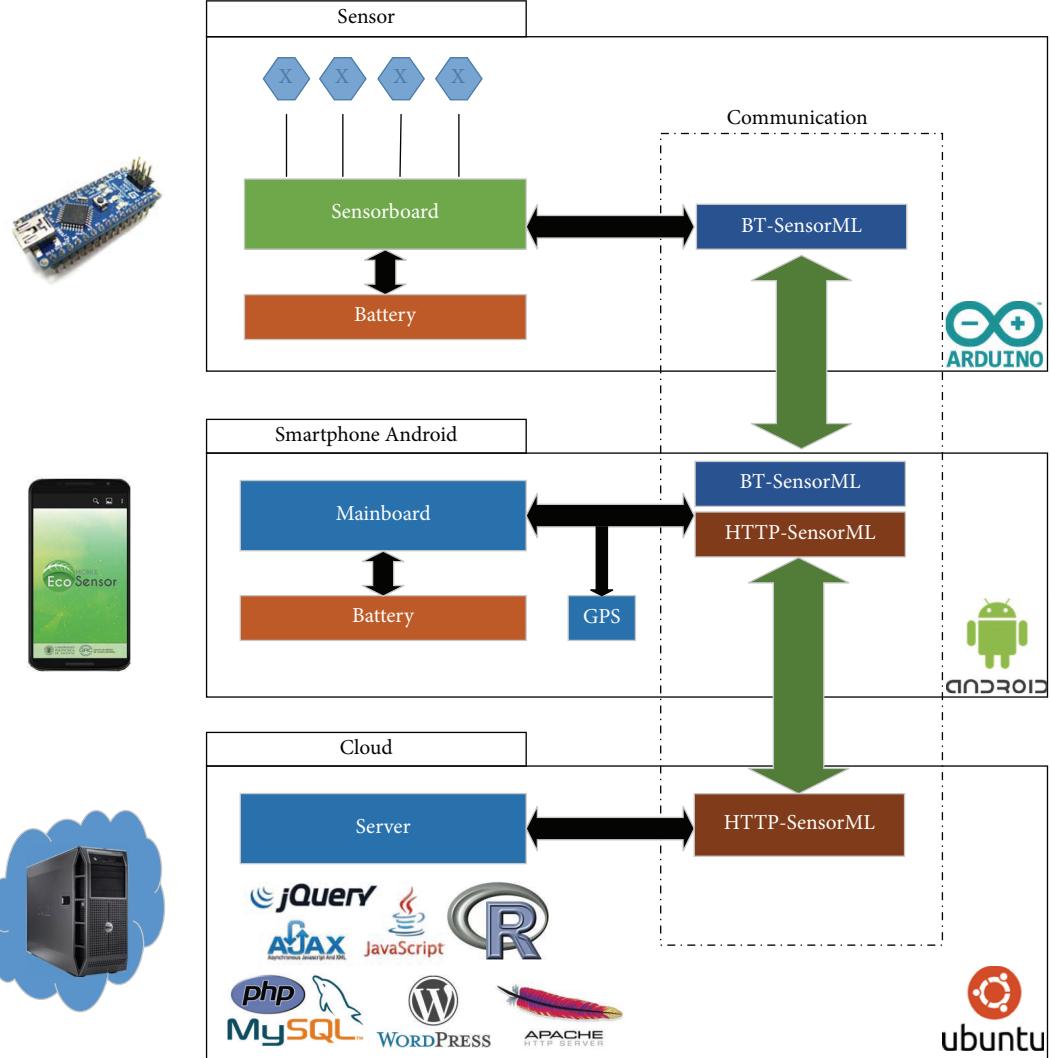


FIGURE 2: Overview of the proposed mobile sensing architecture including the main hardware components and the technologies used.

a central server for data processing, generating detailed reports afterward.

The architecture integrates several hardware and software components. These components are either mobile sensing elements or the central processing server that analyzes collected data and presents detailed information. Mobile sensing elements are composed of two different components: (i) a mobile sensor for measuring pollution data and (ii) an Android-based device for showing real-time pollution status, storing the data, and transferring it to the Cloud server when network connectivity is available. Figure 2 provides an overview of the proposed architecture.

The mobile sensor is based on an Arduino platform [14], and it measures environment parameters through various sensors (ozone, CO₂, air pollution, or temperature). Once data is ready, it can be made available to the Android device via a Bluetooth connection.

For the user to manage the mobile sensing process, an Android application was developed (see Figure 3). This

application allows starting or stopping a trace, view captured data in real-time, uploading data to the server, and perform other management tasks.

Internally, the application has two parts: (i) a service that continually receives the data sent by the sensor and that saves it in an internal database; the service opens a Bluetooth serial communications channel with the sensor for the data transfer; and (ii) a user interface that allows starting or stopping a trace data capture from the sensor and that also provides real-time feedback about pollution levels at the current location according to the AQI index [15]. Moreover, the full trace can be represented on a map showing pollution variations through different color identifiers. Once the trace is completed, the data can be sent to the server via an HTTP connection.

Concerning the Central server, it is a web-enabled system that handles the information received from the Android device. The received data is saved in a MySQL database. Next, the information is processed using different statistical

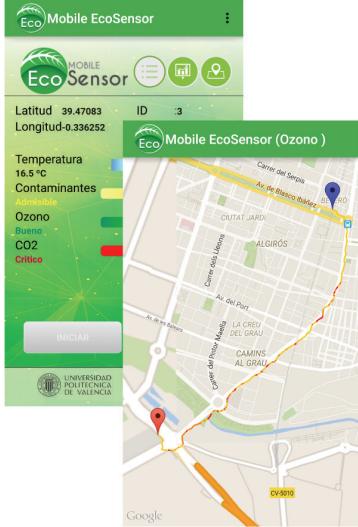


FIGURE 3: Android-based application deployed monitoring screen (back) and path of a monitoring session (front).

procedures. Finally, the detailed information is presented to the system administrator through a web front-end.

The web interface of the Cloud server was built using a Word-press CRM. The website, available at <http://www.ecosensor.net>, allows the administrator to have full access to the information in terms of trace handling, processing, and visualization. Once logged in, the administrator views all uploaded traces and can choose different statistical analyses for the different datasets (e.g., CO₂, ozone, air pollution, and temperature). For statistical analysis and report generation it relies on the R graph tool [16]. The generated graphics for each dataset include heat map, boxplots, time series, and the confidence associated with the spatial interpolation process, as shown in Figure 4.

4. Monitoring Process

After defining the proposed architecture, we now focus on the most relevant issues regarding the reliability of the pollution monitoring process. Our target pollutant was ozone due to its well-known negative impact on health and also because it is more complex to measure accurately than other pollutants due to its dependency on temperature and time of day.

The issues that should be taken into account to perform accurate ozone measurements are the following:

- (i) Sensor output data measurements are highly variable in ranges close to the real values, and so such variability should be reduced.
- (ii) The sensor outputs should be transformed into the respective units for each pollutant. In most cases, the measured resistance value must be converted into *particles per billion* (ppb).
- (iii) In order to use mobile sensors, time-dependent variability must be removed since different samples are obtained at different times.

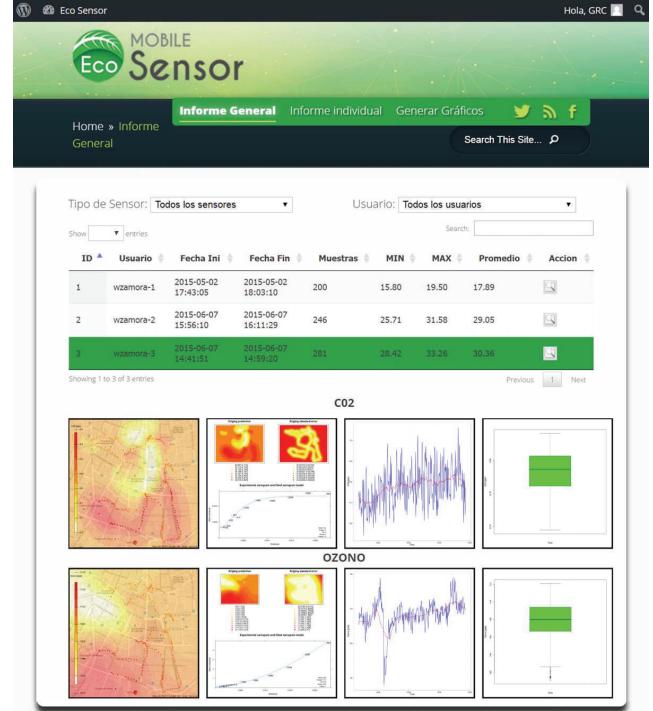


FIGURE 4: Example of the Cloud application web page showing some monitoring sessions and the analysis output for two air pollutants.

- (iv) Using the adjusted measurements, the next phase is to apply spatial interpolation techniques for creating detailed pollution maps.

Figure 5 shows the different steps taken when transforming the raw sensor readings in detailed air pollution maps. Also, we detail bellow how each of these issues has been addressed.

4.1. Data Reading. Low-end sensors introduce significant variability between consecutive measurements (absolute values for intersample differences have $\bar{x} = 6.15$, $\sigma = 5.73$), so data retrieval processes should eliminate these oscillations associated with noise in the sensor readings. For this purpose, we performed the following steps: first, we calculated the average value of 25 samples ($n = 25$), with an interval of 10 ms between each consecutive sample, as shown in

$$O_s = \frac{\sum_{i=1}^n O_i}{n}. \quad (1)$$

In this equation O_s represents the estimated ozone level, O_i represents the ozone level sample i obtained from the sensor, and n represents the number of measurements. In this step, we slightly reduce the absolute variability ($\bar{x} = 5.39$, $\sigma = 5.01$). Afterward, and taking into account that the variability was still very high, we used a low-pass filter for the data analysis process with α equal to 0.95 to further reduce this variability, as shown in

$$O_i = O_r + \alpha \cdot (O_{i-1} - O_r). \quad (2)$$

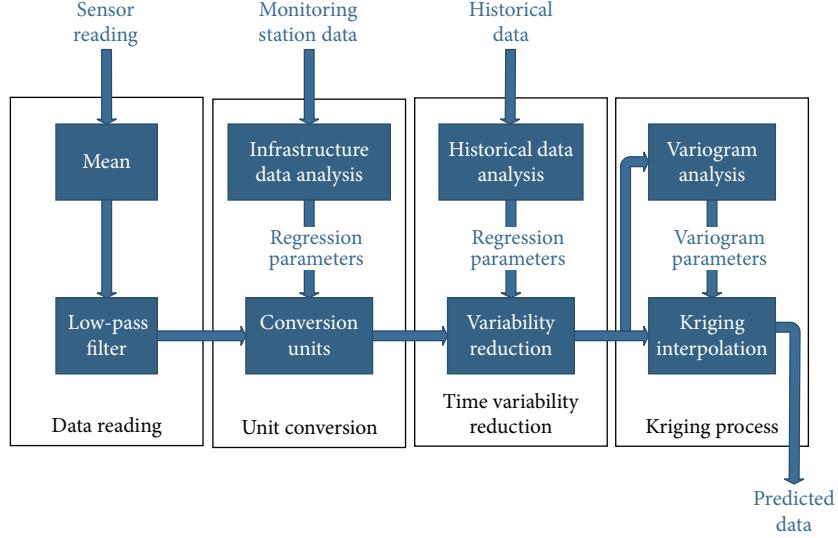


FIGURE 5: Monitoring process overview showing the different tasks associated with each step in the process.

O_i represents the current ozone level, O_{i-1} represents the ozone level in the previous measurement, O_r represents the filtered ozone value, and α represents the filter coefficient. In this step, we drastically reduce the absolute variability ($\bar{x} = 0.32$, $\sigma = 0.30$).

Figure 6(a) shows the difference between the values of captured ozone levels and the values of ozone levels after applying the low-pass filter, and Figure 6(b) shows the variability after applying the mean and the low-pass filter. It shows that data variability is significantly reduced while maintaining the correct trend.

At the end of this process, we have measurements without the variability associated with noisy sampling.

4.2. Unit Conversion. Sensors provide an electrical signal output. It needs to be transformed to a pollution level value. Specifically, the *ozone sensor probe (MiC-2610)* has an internal resistance, which varies proportionally to ozone concentrations. The sensor can measure ozone variations between 10 ppb and 1000 ppb, where that resistance varies between 11 k Ω and 2 M Ω with a quasi-linear behavior.

Sensor specifications were made at a constant temperature of 25 degrees centigrade and vary depending on weather conditions.

For calibrating the sensor we have done several measurements at different days, and under different weather conditions, to get a broad range of values. These data have been compared against the data obtained from the official monitoring station located at the Technical University of Valencia (UPV), Spain. Data obtained are shown in Table 1. Considering that the measurements have a dependency on both ozone levels and temperature, we obtained through regression a second-degree polynomial (see (3)) that takes the temperature and the resistance obtained by the sensor into account to determine the actual ozone values:

$$O = \alpha + \beta_1 t + \beta_2 r + \beta_3 r^2. \quad (3)$$

TABLE 1: Relationship between sensor readings and monitoring station readings.

Resistance [Ohm]	Temperature [°C]	Station ozone [ppb]	Calculate ozone [ppb]
25.0	19.0	80	63.55
31.0	16.0	65	73.10
28.0	15.5	52	55.22
35.0	13.0	83	79.72
28.0	28.0	120	117.46
23.3	23.0	70	77.58
23.5	22.0	70	73.35

In this equation α is a regression coefficient, β_1 is a temperature coefficient, β_2 is a sensor reading coefficient, β_3 is the reading coefficient squared, t is the measured temperature, and r is the sensor reading (measured as resistance). The output O is the ozone level measured. The final regression obtained is shown in

$$O = -29.19 + 4.79t - 3.09r - 0.13r^2. \quad (4)$$

The error obtained for the regression was $R^2 = 0.85$. Compared against a 1st-order regression ($R^2 = 0.83$) the obtained result is better in terms of R^2 . Compared against a 3rd-order regression ($R^2 = 0.86$), the improvement in R^2 is minimum and differences are minor.

4.3. Time Variability Reduction. To cover large areas of land with a fine spatial granularity we use mobile sensors, which can capture data at various points although at different time instants. So, the difference between measurements O has both time ΔO_t and spatial ΔO_e dependencies. Since our main

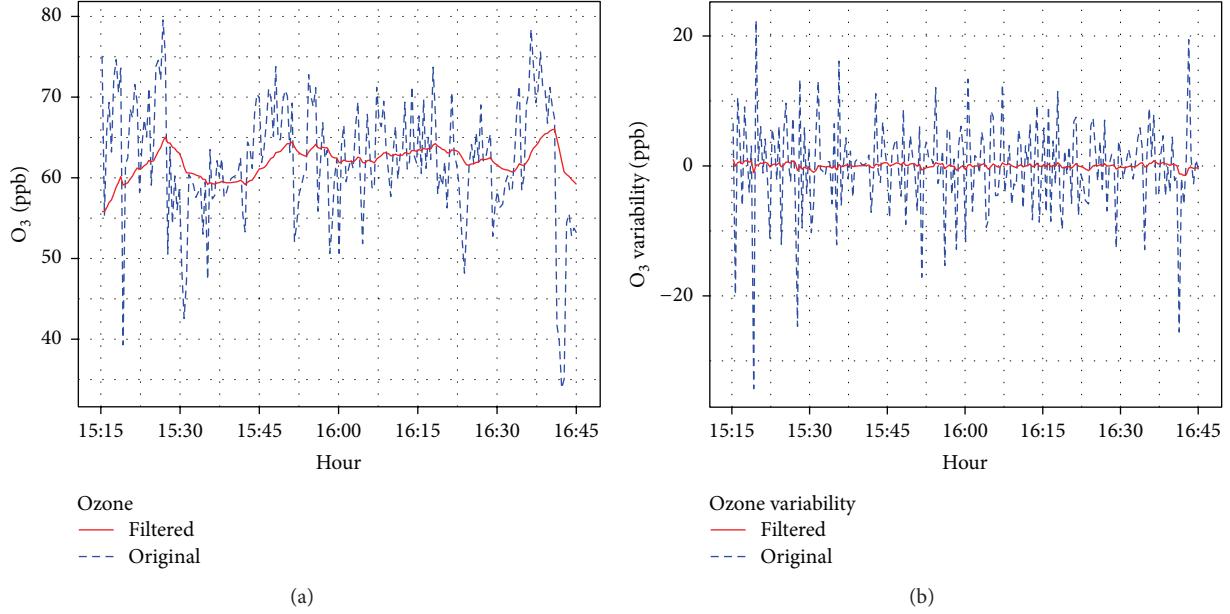


FIGURE 6: Relationship between captured data and filtered data (a) and relationship between captured data variation and the filtered data variation (b).

goal is to determine differences between ozone levels in a particular area, it is necessary to eliminate the time variation:

$$\begin{aligned}\Delta O &= \Delta O_t + \Delta O_e, \\ \Delta O_e &= \Delta O - \Delta O_t.\end{aligned}\quad (5)$$

For the calculation of the ozone time variations, we analyzed data from a monitoring station located at the Technical University of Valencia, focusing on historical data between 2008 and 2014. In the historical data analysis, we analyzed the ozone evolution focusing on average monthly measurements between 2008 and 2014. It is noted that the values are higher from April to September and lower for the remaining months. Figure 7 shows the mean values and standard deviation in the shaded area and maximum values with the top line. The variation in ozone levels during a representative June day was also analyzed. As shown in Figure 8, ozone levels reach their lowest value at the end of the night, at about 6 am, and rise to reach maximum values at 2 or 3 p.m., beginning to decline gradually afterward. The behavior for the other months of the year is analogous to the month shown.

As a result of the analysis of these data, we observe that ozone has a different behavior in summer (specifically from April to September) compared to the rest of the year. During day time, the behavior is very similar to the parabolic logarithmic distribution, with an onset of rapid growth followed by a less pronounced decline.

Based on the previous data regarding monthly average values between 2008 and 2014, taken at the monitoring station of the Technical University of Valencia, ozone level prediction relies on a parabolic logarithmic regression influenced by temperature and season of the year, one for summer, and one

for winter. The expression used (in linear format) was the following:

$$\ln(O_t) = \alpha + \beta_1 s + \beta_2 t + \beta_3 \ln(h) + \beta_4 \ln(h)^2, \quad (6)$$

where h is time of day, s is the season, t is the temperature, and the remaining α and β_i values are regression coefficients (β_1 : season coefficient, β_2 : temperature coefficient, β_3 : coefficient for the logarithm of the time of day, and β_4 : coefficient for the logarithm of the time of day squared):

$$\begin{aligned}\ln(O_t) &= -7.70 + 0.03s - 0.01t + 9.23 \ln(h) \\ &\quad - 1.77 \ln(h)^2, \\ \ln(O_t) &= -15.43 + 0.12s + 0.03t + 14.42 \ln(h) \\ &\quad - 2.83 \ln(h)^2.\end{aligned}\quad (7)$$

The values of $\|R^2\|$ are 0.91 and 0.82 for summer and winter, respectively, showing a behavior very similar to the actual one.

The procedure followed to correct time-dependent variability was as follows: (i) ozone values are calculated at two time instants using (6); (ii) the difference between the values is obtained; and (iii) the actual readings are reduced according to the calculated variation.

4.4. Interpolation Data. The adjusted data is the input for creating detailed pollution maps. In the scope of this work this is achieved by using the R graph tool. Specifically, we rely on spatial interpolation techniques known as ordinary kriging. First, a semivariogram is calculated for a specific area, and kriging parameters are determined. Next, a detailed

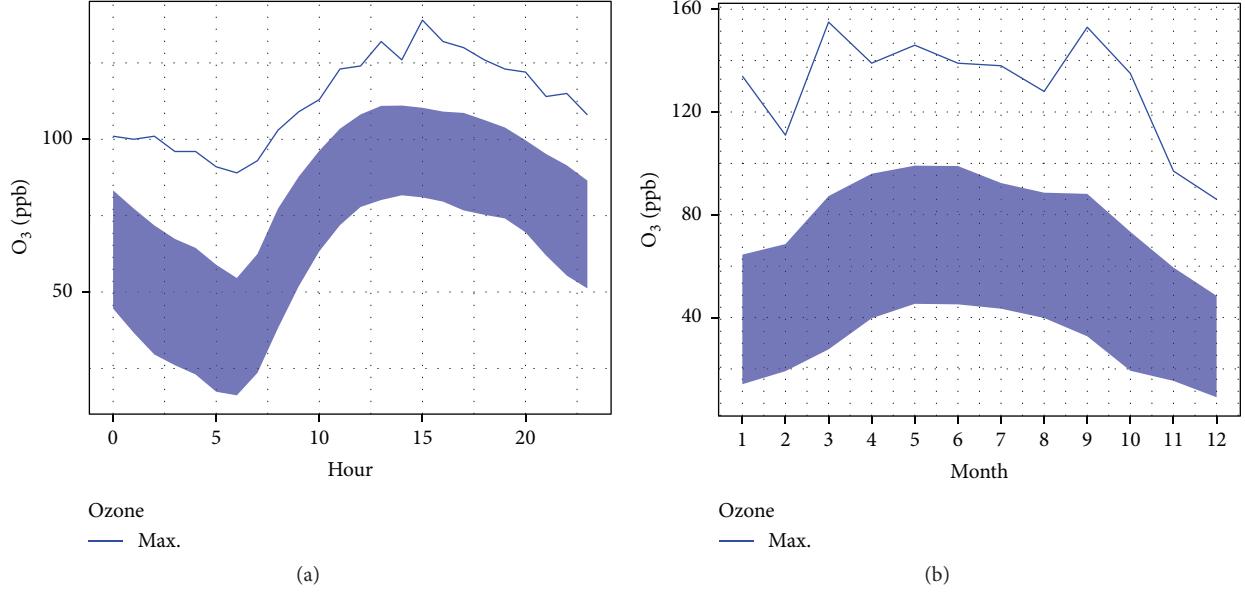


FIGURE 7: Ozone evolution in June (a) and throughout the year (b).

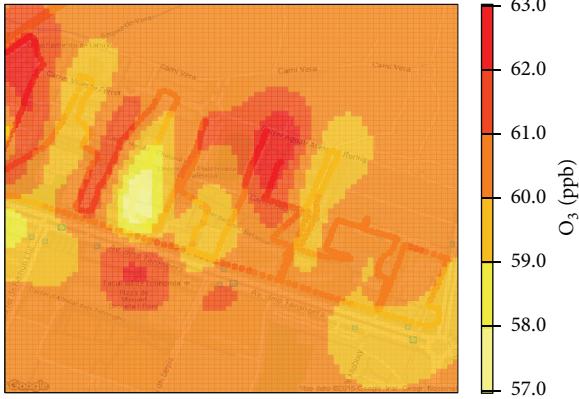


FIGURE 8: Example of an ozone distribution heat map for the UPV using the proposed architecture.

pollution distribution is created using the obtained parameters. To easily visualize the pollution levels distribution in space, different maps are created, as shown Figure 8.

The semivariogram defines the variance of the differences between two points. It determines the parameters required for the kriging interpolation, which have an influence on the distribution form.

- (i) Sill determines the total variance of the values.
- (ii) Nugget determines the variance at the origin.
- (iii) Range determines the range of influence of the model.
- (iv) Model determines the distribution function. It can be Gaussian, Spheric, Exponential, Circular, or Linear.

Figure 9 shows a sample semivariogram as an example.

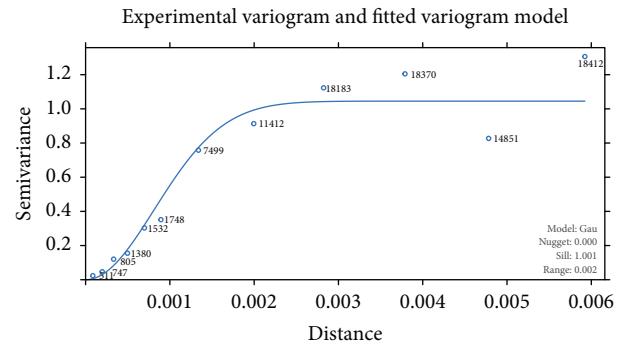


FIGURE 9: Example of a semivariogram showing a Gaussian distribution. It shows the different parameters related to the interpolation techniques (Nugget, Sill, and Range).

5. Finding the Optimal Measurement Strategy

After defining the architecture and the monitoring process, we now proceed to determine the optimal strategy for air pollution data collection using mobile sensors.

With this purpose, we first analyzed the impact of mobility on sensor readings by comparing static against mobile measurements. Also, we determined the influence of sensor orientation in the mobile sensing process. Our next step was to analyze the impact of reducing the sampling frequency on the kriging process accuracy under mobile scenarios. Similarly, we analyzed the impact of reducing the number of spatial samples on the kriging process accuracy. This was achieved by skipping selected streets when capturing data, progressively reducing the overall path.

5.1. Optimal Sensor Positioning. To analyze the impact of mobility on the data capture process we performed different

TABLE 2: Statistical summary of the sensor position analysis.

Period	Mean	Std. Dev.	<i>p</i> value
Static	27.39	0.85	—
Movement	27.34	1.03	0.25
Facing forward	27.41	1.04	0.77
Facing backwards	27.44	1.02	0.38
Facing upwards	26.85	0.95	0.06

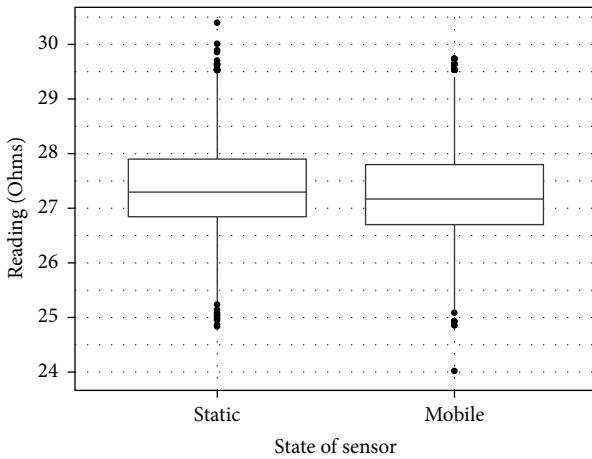


FIGURE 10: Analysis of the variability of mobile sensor readings: static versus mobile sensor.

tests, collecting ozone levels in a specific area either statically or using a bike moving at a speed of about 20 km/h. For mobility tests, we collected measurements with different sensor orientations: (i) facing forward, (ii) facing backwards, and (iii) facing up. Statistics for the “mobile” case combine measurements with different sensor orientations.

To have further insight into how these results are distributed, Figure 10 shows that mobility, at least at the speed used for testing, does not have a significant impact on sensor measurements.

The results for the *t*-test analysis are shown in Table 2, revealing that we cannot find a statistically relevant difference between the static sensor ($\bar{x} = 27.39$, $\sigma = 0.85$) and the mobile sensor ($\bar{x} = 27.34$, $\sigma = 1.03$), obtaining a *p* value = 0.25 with a $\alpha = 0.05$, neither for the facing forward orientation ($\bar{x} = 24.41$, $\sigma = 1.04$, *p* value = 0.77) nor for the facing backwards orientation ($\bar{x} = 27.44$, $\sigma = 1.02$, *p* value = 0.38).

Figure 11 shows that the actual sensor orientation has little impact on the data capture process, keeping the differences between different orientations minimal. Anyway, the backwards orientation option shows greater resemblance with the static measurements and was adopted for the tests that follow.

5.2. Impact of Time Sampling on Geostatistical Predictions. In this section we analyze the impact of time sampling on the predicted pollution map. In particular, we want to determine if reducing the number of samples allows making similar predictions or if, on the contrary, there is a significant

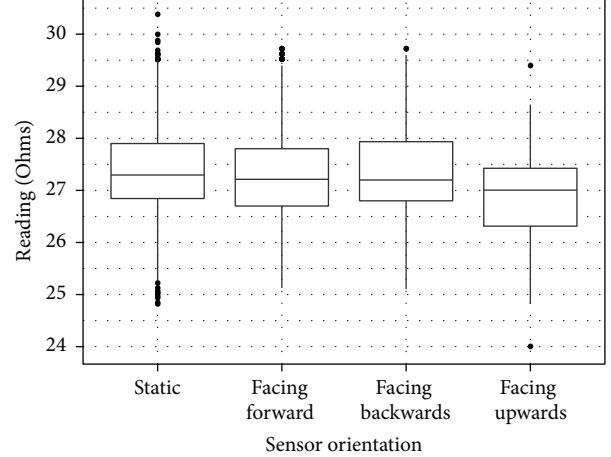


FIGURE 11: Analysis of the variability of mobile sensing for different sensor orientations.

TABLE 3: Statistical summary of the time sampling analysis.

Period	Mean	Std. Dev.	Similarity (s_i)
5 sec.	60.31251	1.140371	1
10 sec.	60.31928	1.158987	0.9734
20 sec.	60.37815	1.131514	0.9579
30 sec.	60.48890	1.118012	0.9411
40 sec.	60.36123	1.131782	0.9225
80 sec.	60.45629	1.126616	0.9181

prediction error when generating the pollution map. For this purpose, we monitored the Technical University of Valencia campus with a mobile ozone sensor installed on a bike.

To obtain an accurate distribution of ozone levels, we monitored the entire campus by setting the sampling period to the lowest value allowed by the sensor (5 seconds). Next, we reduced the sampling frequency by setting the intersample period to 10, 20, 30, 40, and 80 seconds. This was achieved by filtering the full trace and retrieving datasets with 1/2, 1/4, 1/6, 1/8, and 1/16 of the data, respectively.

Next, we performed spatial interpolation through kriging for each trace, obtaining a detailed pollution distribution. We used the full trace (samples every 5 seconds) as reference and compared it against the results obtained using the other datasets.

Table 3 summarizes the statistical analysis for the different datasets in terms of mean, standard deviation, and relative prediction error, with the latter being calculated using the initial trace (5 s sampling) as reference, as shown in

$$s_i = 1 - \frac{1}{m \cdot n} \sum_{x=0}^m \sum_{y=0}^n \left| \frac{k_{i_{xy}} - k_{0_{xy}}}{\Delta k_0} \right|. \quad (8)$$

In this equation, s_i represents the similarity index of dataset i with respect to the reference dataset, m and n represent the width and length of the target area under analysis, $k_{i_{xy}}$ represents the value calculated through kriging interpolation for dataset i at position xy , $k_{0_{xy}}$ represents

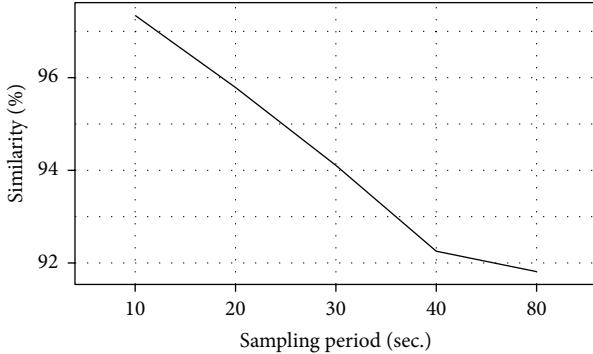


FIGURE 12: Analysis of the output similarity with respect to reference sampling period (5 s).

the value calculated through kriging interpolation for the reference dataset at position xy , and Δk_0 represents the total variation of the predicted values for the reference dataset.

By analyzing Table 3 we can see that the mean and the standard deviation values are nearly the same in all cases, although the similarity index s_i varies more significantly. This information is also shown in Figure 12 for the sake of clarity. Notice that, although the distribution of values is similar, the mean similarity shows an almost linear decrease. Nevertheless, the similarity values are still relatively high since the kriging interpolation process also acts as an error filter, helping to approximate the mean value when lacking enough reference values.

Detailed heat maps for some relevant traces (5 seconds, 20 seconds, and 80 seconds) are shown in Figure 13. By taking a look at these heat maps, built through the kriging interpolation process, we can clearly see that the level of detail experiences a degradation. In particular we find that, although the pollution maps for intersample times of 5 seconds and 20 seconds are quite similar, significant differences are observed when the sampling period grows to 80 seconds; for the latter case, the ozone distribution achieved is quite different from the one used as reference (5 seconds). Based on these maps, it becomes quite clear that little differences in terms of basic statistical analysis can represent huge differences in terms of the spatial distribution of those values.

5.3. Impact of Spatial Sampling on Geostatistical Predictions. In this section we analyze the impact of spatial sampling on the predicted pollution map. In particular, we want to determine to which degree taking a shorter, less exhaustive path throughout the target area (reducing the trip time and the number of samples accordingly) affects the accuracy of the predictions made.

To find the optimal spatial sampling strategy we produce different datasets by deleting path fragments from the initial trace. In detail, starting from the full trace (100% of the data), we deleted selected paths so as to produce shorter but yet valid trips, maintaining start and end locations. As a result, we obtained traces with 72%, 54%, 50%, 46%, and 42% of the data.

TABLE 4: Statistical summary of the spatial sampling analysis.

Dataset size	Mean	Std. Dev.	Similarity (s_i)
100%	60.31251	1.140371	1
72%	60.49253	1.000335	0.9336
54%	60.62813	1.112316	0.9084
50%	60.66518	1.137273	0.8820
46%	60.66079	1.137295	0.8872
42%	60.51269	1.082692	0.8530

Similar to the previous section we perform, for each dataset, a statistical analysis of the resulting data, also obtaining the pollution heat map generated through kriging interpolation and calculating the similarity index using (8).

Table 4 presents the statistical analysis results showing the mean, the standard deviation, and the similarity, with the latter being calculated using the initial dataset as reference.

Based on Table 4, we find that the mean value is close to the reference one (60.31) in all cases, although being in general slightly higher. This occurs because the first eliminated path showed the lowest values.

Figure 14 shows the decreasing trend when spatial sampling decreases. Compared to the time sampling results of Figure 12, we find that now the similarity values degrade much faster, meaning that reducing the route taken along the target area is prone to eliminate relevant samples, resulting in a less detailed pollution map.

Figure 15 shows detailed maps for datasets representing 100%, 72%, 50%, and 42% of the data. Based on these heat maps, we can see clearly how spatial subsampling causes a distortion on the spatial distribution of pollution throughout the target area.

Overall, we can conclude that the spatial sampling granularity is the most relevant factor to take into account, with time sampling granularity being less but yet somehow important, and sensor orientation the factor having less impact on results.

6. Validation of the Proposed Approach

As stated at the beginning of the paper, the current infrastructure elements allow measuring pollution levels in cities with high accuracy, although with a low spatial resolution. On the contrary, our proposed mobile sensing approach is able to achieve a much higher spatial resolution using cheap sensors. Thus, in this section, we validate our approach by first comparing captured values with the range of values typical of the time of year and then by comparing the ozone maps generated when relying on either infrastructure-based or mobile-based sensing.

We started by gathering data in different areas of Valencia using the proposed mobile sensors. Different experiments have been conducted at different times, allowing us to compare the data captured with the data from the existing public infrastructure. In particular, for each route taken, we first reduced the data variability using the proposed low-pass filter (see (2)). Next, the measurements were adjusted through (3).

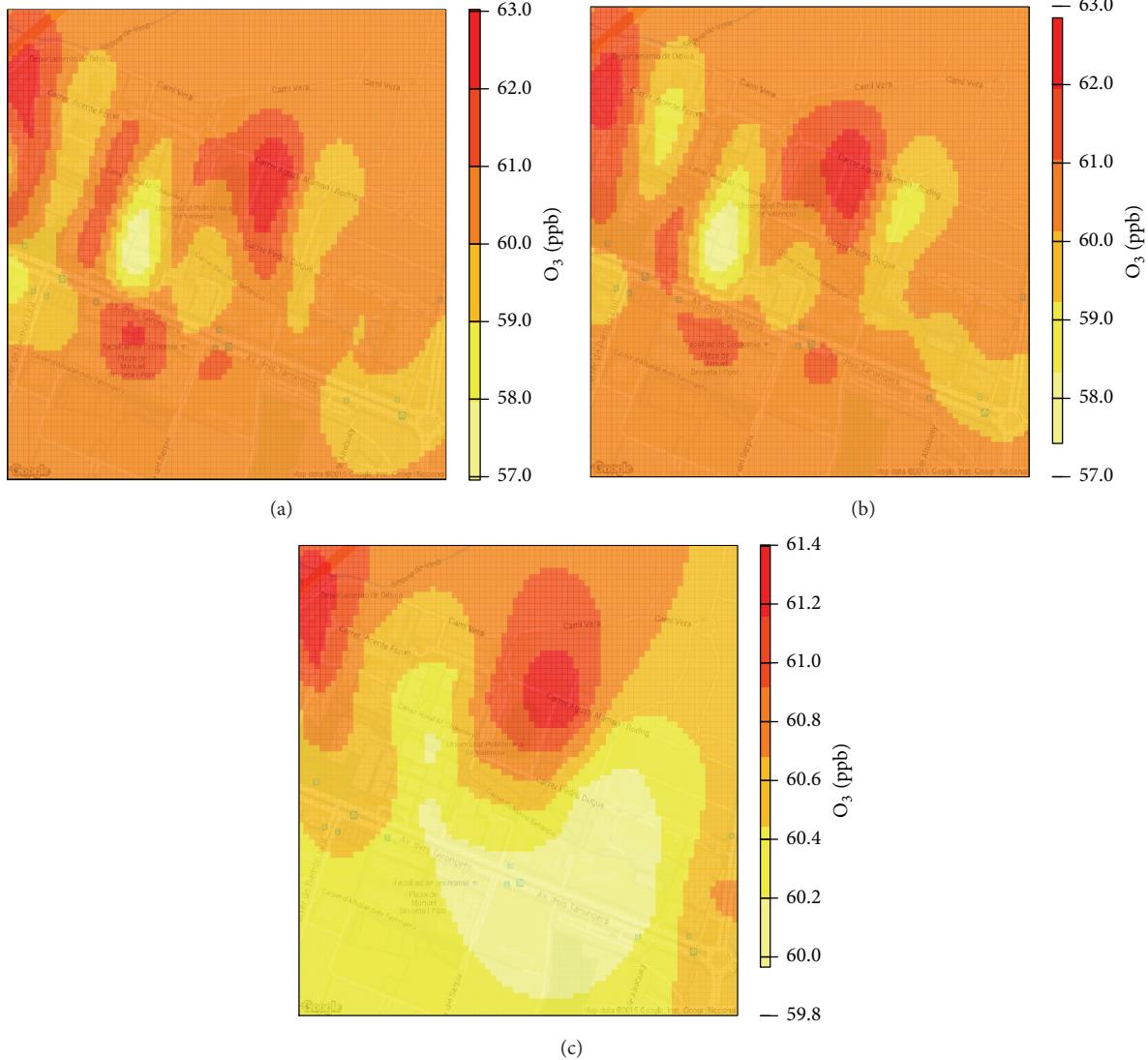


FIGURE 13: Heat maps for the ozone distribution using different sampling periods (5, 20, and 80 seconds).

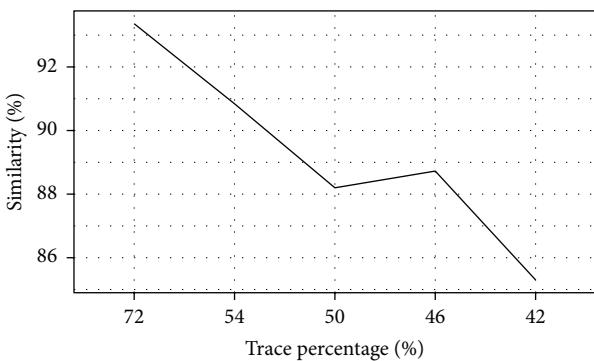


FIGURE 14: Analysis of the similarity with respect to the reference trace (100% of the data used).

Finally, the temporal dependencies of data were reduced according to (6).

Figure 16 shows data for a particular route and the common values at the date of the capture (February 16, 2015). We can see that the measured ozone levels are within the range of historical values for the monitored time, being quite close to the expected value (mean). This indicates that, using our methodology, we are able to obtain reliable data despite using low-cost sensors, allowing us to focus our analysis on the spatial variations of pollutants.

We now proceed to compare the actual heat maps for a specific date and time of day using only infrastructure data and only data obtained by our sensor. We can see that, by relying on our proposed architecture (see Figure 18), it becomes possible to observe in detail even small pollution variations, while using only infrastructure-based data (see Figure 17) the observed variations are much smoother, experiencing a linear increase or decay from one air quality station to the other.

Overall, it becomes clear that, despite having up to 5 different stationary air quality stations in the city of Valencia, they fail to capture significant details that are related to

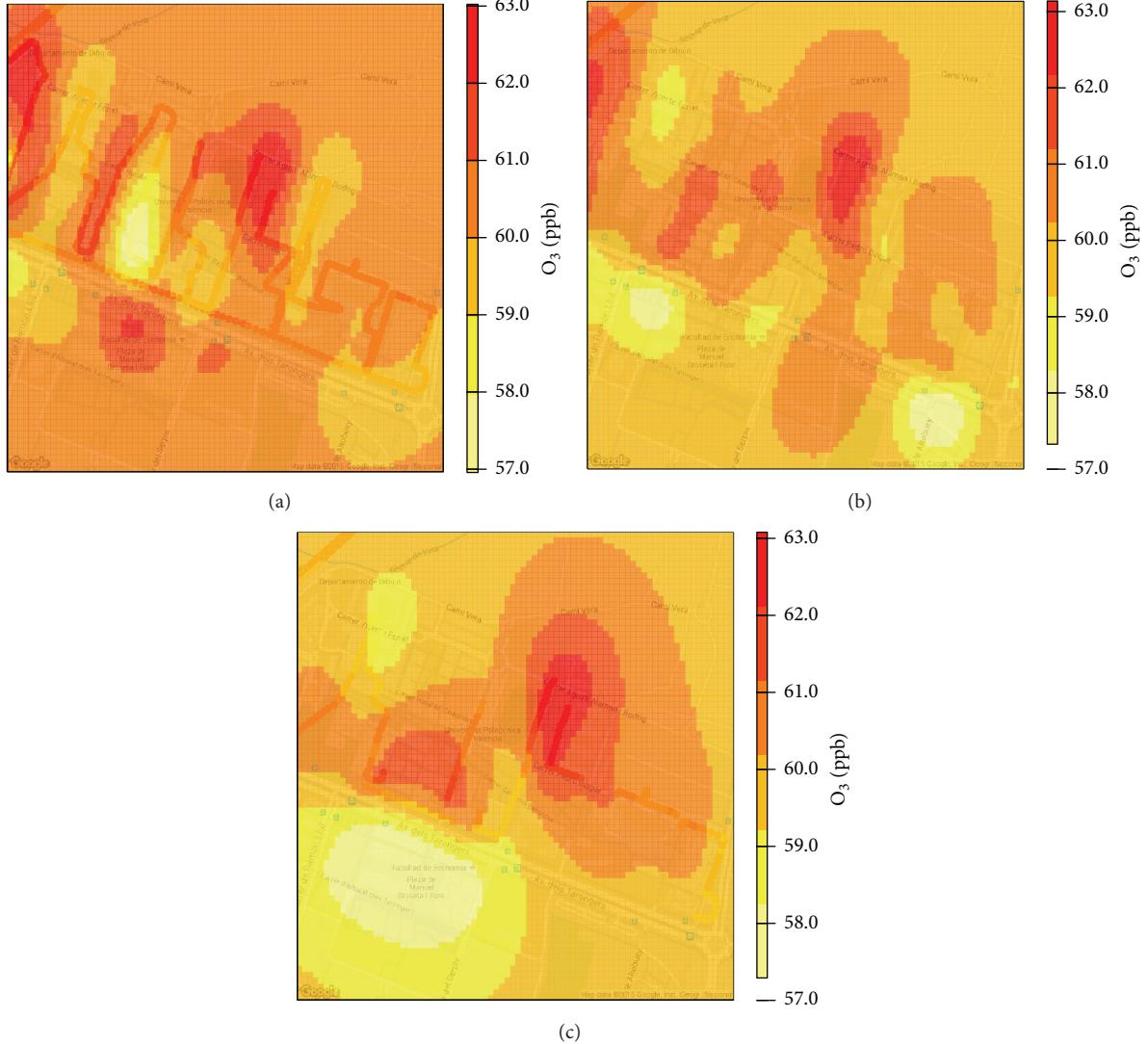


FIGURE 15: Heat maps for the ozone distribution using different fraction of the original trace (100%, 72%, and 42%).

areas with more traffic congestion (high pollution values) or green/windy spaces (low pollution values), thereby leading to some wrong conclusions. In contrast, our approach is able to provide a greater richness since all small variations can be perceived with great detail, thereby meeting the proposed goal.

7. Conclusions and Future Work

Nowadays, environment pollution monitoring has become a fundamental requirement for cities worldwide, and there are many studies related to it. Nevertheless, only a few explore all sides of this problem.

In this paper we proposed a complete architecture for environmental monitoring that combines low-end sensors, smartphones, and Cloud services to measure pollution levels with a high spatial granularity. In detail, we used a mobile sensor to provide pollution measurements, a smartphone

providing real-time feedback about air quality conditions and also acting as a gateway by uploading gathered data to the Cloud server, in addition to the Cloud server itself, required for data processing and visualization.

Once the architecture was defined, we analyzed different issues related to the monitoring process: (i) filtering captured data to reduce the variability of consecutive measurements; (ii) converting the sensor output to actual pollution levels; (iii) reducing the temporal variations produced by the mobile sensing process; and (iv) applying interpolation techniques for creating detailed pollution maps.

To address the challenges associated with taking mobile measurements in a target area, we analyzed the influence of the sensor orientation in the data capture process, as well as the impact of time and spatial sampling. In particular, we varied the sampling period and the overall path length to determine the most effective monitoring strategy. Experimental results show that the sensor orientation and the sampling period, within certain bounds, have very little

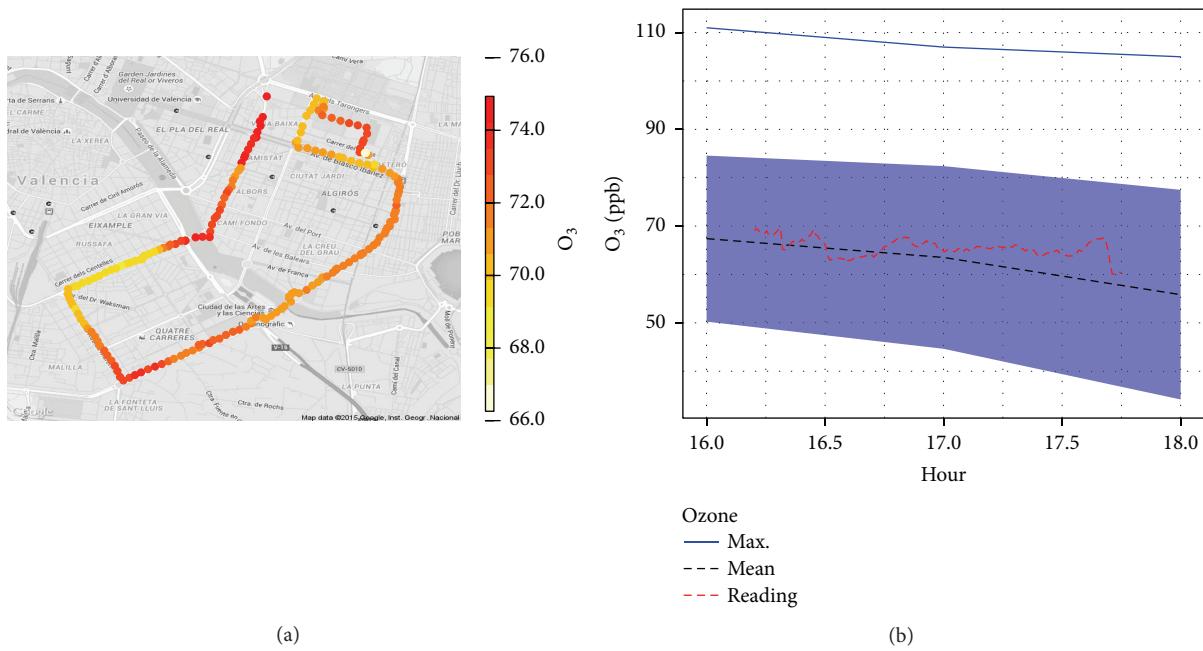


FIGURE 16: Comparison between captured data and typical values for the day/time of the year.

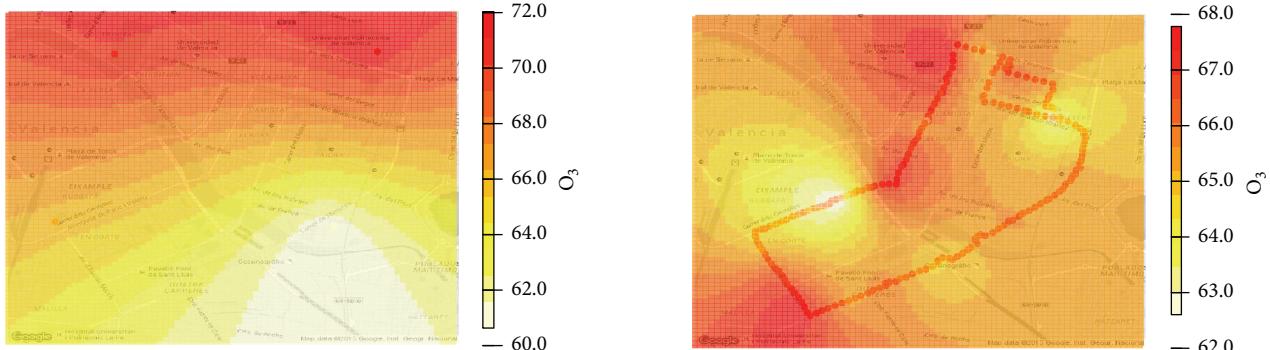


FIGURE 17: Ozone levels in the target region using infrastructure data.

influence on the data captured, while the actual path taken has a greater impact on results, especially when estimating the distribution of pollutants throughout the target area.

Finally, we validated our proposal by comparing values obtained by our mobile sensor with typical values from monitoring stations at the same dates and location. Furthermore, we compared the resulting heat maps generated using data from monitoring stations against ours, showing that using our mobile sensing approach is able to provide a much higher data granularity.

The next steps in this research include improving the spatial interpolation process and comparing different sensor types.

Competing Interests

The authors declare that they have no competing interests.

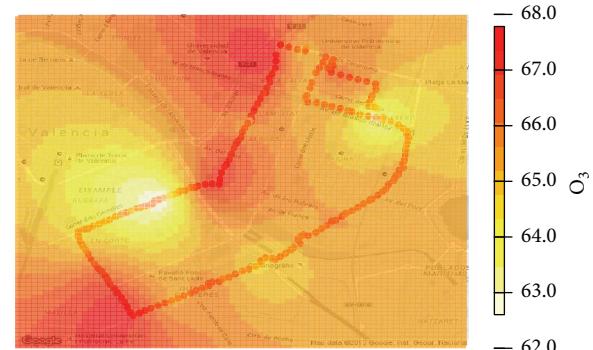


FIGURE 18: Ozone level in the target region using mobile sensing data.

Acknowledgments

This work was partially supported by the “Programa Estatal de Investigación, Desarrollo e Innovación Orientada a Retos de la Sociedad, Proyecto I+D+I TEC2014-52690-R,” the “Universidad Laica Eloy Alfaro de Manabí,” and the “Programa de Becas SENESCYT de la República del Ecuador.”

References

- [1] S. Król, B. Zabiegała, and J. Namieśnik, “Monitoring VOCs in atmospheric air I. On-line gas analyzers,” *TrAC—Trends in Analytical Chemistry*, vol. 29, no. 9, pp. 1092–1100, 2010.
- [2] R. E. Basher, “Review of the Dobson spectrophotometer and its accuracy,” in *Atmospheric Ozone*, C. S. Zerefos and A. Ghazi, Eds., pp. 387–391, Springer, New York, NY, USA, 1985.
- [3] J. Fishman, V. Ramanathan, P. J. Crutzen, and S. C. Liu, “Tropospheric ozone and climate,” *Nature*, vol. 282, no. 5741, pp. 818–820, 1979.

- [4] T.-M. Chen, J. Gokhale, S. Shofer, and W. G. Kuschner, "Outdoor air pollution: ozone health effects," *The American Journal of the Medical Sciences*, vol. 333, no. 4, pp. 244–248, 2007.
- [5] T.-M. Chen, J. Gokhale, S. Shofer, and W. G. Kuschner, "Outdoor air pollution: particulate matter health effects," *The American Journal of the Medical Sciences*, vol. 333, no. 4, pp. 235–243, 2007.
- [6] R. D. Brook, J. R. Brook, B. Urch, R. Vincent, S. Rajagopalan, and F. Silverman, "Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults," *Circulation*, vol. 105, no. 13, pp. 1534–1536, 2002.
- [7] A. Adam-Poupart, A. Brand, M. Fournier, M. Jerrett, and A. Smargiassi, "Spatiotemporal modeling of ozone levels in Quebec (Canada): a comparison of kriging, land-use regression (LUR), and combined bayesian maximum entropy—LUR approaches," *Environmental Health Perspectives*, vol. 122, no. 9, pp. 970–976, 2014.
- [8] L.-J. S. Liu and A. J. Rossini, "Use of kriging models to predict 12-hour mean ozone concentrations in metropolitan Toronto—a pilot study," *Environment International*, vol. 22, no. 6, pp. 677–692, 1996.
- [9] M. Brković and V. Sretović, "Urban sensing-smart solutions for monitoring environmental quality: case studies from Serbia," in *Proceedings of the 48th ISOCARP World Congress (ISOCARP '12)*, 2012, <http://bit.ly/Z5AarU>.
- [10] H. B. Deng and L. Zhang, "Design on ZigBee wireless sensor network node," *Key Engineering Materials*, vol. 474–476, pp. 283–286, 2011.
- [11] C. T. Calafate and B. Durocourtial, "On the use of mobile sensors for estimating city-wide pollution levels," in *Proceedings of the International Wireless Communications and Mobile Computing Conference (IWCMC '15)*, pp. 262–267, IEEE, Dubrovnik, Croatia, August 2015.
- [12] Y. Cheng, X. Li, Z. Li et al., "AirCloud: a cloud-based air-quality monitoring system for everyone," in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems (SenSys '14)*, pp. 251–265, Memphis, TN, USA, November 2014.
- [13] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: when urban air quality inference meets big data," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13)*, pp. 1436–1444, August 2013.
- [14] Arduino, 2015, <https://www.arduino.cc/>.
- [15] United States Environmental Protection Agency, *Air Quality Index*, United States Environmental Protection Agency, 2015, <http://cfpub.epa.gov/airnow/index.cfm?action=aqibasics.aqi>.
- [16] R-Foundation, *R Project*, 2015, <https://www.r-project.org>.

