

EcoSensor: Monitoring environmental pollution using mobile sensors

Óscar Alvear, Willian Zamora, Carlos T. Calafate, Juan-Carlos Cano and Pietro Manzoni

Department of Computer Engineering

Universitat Politècnica de València,

Camino de Vera, S/N 46022 Valencia, Spain, +34.963877007

Email: osal@doctor.upv.es, wilzame@posgrado.upv.es, {calafate,jucano,manzoni}@disca.upv.es

Abstract—Air pollution monitoring has become an essential requirement for cities worldwide. Currently, the most extended way to monitor air pollution is via fixed monitoring stations, which are expensive and hard to install. To solve this problem, we have developed EcoSensor, a solution to monitor air pollution through mobile sensors. It is deployed with off-the-shelf hardware such as Waspote (based on the Arduino platform), low-end sensors, and Raspberry Pi devices.

EcoSensor collects air pollution using embedded sensors and transfers the captured data to an Android-based device, which displays to the user the air pollution levels in real time. EcoSensor also stores the different pollution traces to a Cloud-based server to analyze the pollution distribution. The cloud server uses the uploaded data, together with highly-accurate data made available by the existing air monitoring infrastructure, to create detailed pollution distribution maps using kriging-based spatial prediction techniques.

To optimize the usage of our system, we analyze the impact of sensor orientation in the presence of mobility. Also, we analyze the best time and space sampling strategies to determine the most effective data capturing strategy. Experimental results show that the sensor orientation and the sampling period have a lot less impact on created maps than the actual path taken.

Index Terms—Air Pollution Monitoring, Mobile Sensing Architecture, Geospatial interpolation.

I. INTRODUCTION

AIR pollution is a critical aspect to consider nowadays because it directly impacts people's lives specially in crowded cities, causing health problems (mainly in the respiratory tracts), provoking climate changes, and reducing crop production, among others. Pollution consists in the emission of gases or particles into the atmosphere, producing changes in its composition. Air pollution can be of two types: (i) primary air pollution, when gases or particles are emitted directly into the atmosphere, like carbon monoxide (CO), carbon dioxide (CO₂), particulate matter smaller than 10 microns (PM₁₀), or particulate matter smaller than 2.5 microns (PM_{2.5}); and (ii) secondary air pollution, when gases are produced by a chemical reaction between primary pollutants and some environment element, like ozone (O₃), which is produced by the combination of nitrogen oxides (NO_x), Oxygen(O₂), Volatile Organic Compounds (VOC), and sunlight.

The human factor is the leading cause of air pollution. In particular, the main causes are industrial processes that use fossil fuels, or the automotive park in large cities, both emitting a significant amount of pollutants to the atmosphere.

To have a strict control of air pollution, frequent measurements should take place at all representative locations. For this purpose, there are several institutions worldwide dedicated to the control and monitoring of pollution levels, such as the United States Environment Protection Agency (EPA) [1], which tracks the evolution of air pollution, and the European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM) [2], which brings together 14 European organizations for the analysis and monitoring of climate change.

Most European cities deployed static sensors to monitor and control the evolution of pollution levels on a large scale, providing coarse-granularity pollution levels for a city. In all Europe there are about one thousand five hundred air pollution monitoring stations [2]. In the studied area, Valencia (Spain), there are only 5 stations. However, to measure pollution changes in a city with better granularity, it would be necessary to have many of these monitoring stations, which becomes unfeasible and highly expensive.

Traditional equipment for measuring environmental pollution is typically installed on a static monitoring station. In these stations, the monitoring relies on sophisticated sensors, which are very accurate and introduce minimum oscillations to the data capture process. 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 vehicles to monitor different parts of a city concurrently and continuously. 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.

In this paper we propose EcoSensor, a system to monitor air pollution in crowded areas through low-end sensors. They could be installed in the public transport system or bicycles. The sensor collects pollution data and transfers such data to a Cloud system via a smartphone. In the Cloud server, our system analyzes the pollution distribution through spatial

interpolation techniques to generate detailed pollution maps of the city.

This paper is organized as follows: Section II reviews the related work. Section III describes the architecture and schema of our proposal. Section IV describes our developed EcoSensor applications. In Section V, we analyze the optimal time and space sampling, as well as the impact of sensor orientation in the presence of mobility. Finally, Section VI presents the conclusions and future work.

II. RELATED WORK

In the last twenty years, air pollution monitoring has become very important worldwide due to the influence of air quality in our lives. There are many research works that study the effects of air pollution on our health. Among them we can find the contributions of Chen et al. [4], [5], who analyzed the effects of ozone and particulate 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. In this regard, studies such as [7] and [8] have relied on kriging interpolation techniques to predict pollution 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. In the literature we can find several works adopting this approach. For instance, Brković et al. [9] propose a system to monitor environment pollution in the city of Belgrade using Wasmote sensors installed in the public transport system. Hu et al. [10] use a vehicular sensor network for air pollution monitoring. In particular, they propose to use taxis for deploying the system, analyzing the communication between them. More recently, Calafate and Ducourthial [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. The authors evidence how these techniques allow inferring environmental pollution levels with better granularity.

Our proposal differs from the former ones because it aims at providing an integral solution. In particular, we combine low-end sensors, smart-phones, and cloud services to efficiently monitor pollution levels. By relying on the data readings provided by the existent infrastructure (high reliability), we are

able to calibrate and adjust data readings coming from a wide set of mobile pollution sensors. Then, over the obtained data, we apply kriging techniques to perform spatial interpolation, thereby obtaining detailed pollution maps.

III. ECOSENSOR

EcoSensor is a system that allows monitoring air pollution in a simple and cheap manner, being specially useful in very crowded cities. It combines data coming from existing air quality monitoring stations with data collected by mobile sensors to generate fine-grained reports about pollution levels. Mobile sensors can be installed in bicycles or along the public transportation system to monitor the city in a simple and effective manner. All collected information is stored on a central cloud server for data processing, generating detailed reports afterward.

The architecture of the proposed system integrates several hardware and software components. These components can be classified as mobile sensing elements or cloud elements, being the latter a set of services running on a server that analyze collected data and present detailed information. Mobile sensing elements are composed by three different components: (i) a Wasmote sensor for measuring pollution data, (ii) a Raspberry Pi that acts as a gateway between the sensor and the Android-based device, and (iii) 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. This architecture is shown in Figure 1.

The Wasmote sensor [14] is based on an Arduino platform [15], and it measures air quality through various sensors (Ozone, CO₂, Air Pollution, and temperature). Moreover, it has a GPS interface that allows determining the exact location of each measurement. Once data is ready, it is transferred to the Raspberry Pi [16] via Zigbee [17].

The Raspberry Pi [16] acts as a gateway between the Wasmote sensor and the Android-based device. It has a Raspbian operating system, and it is programmed in Python. It has two communication interfaces: ZigBee [17] for connecting to the Arduino platform, and Bluetooth [18] for transferring data to the Android device.

The Android-based device shows, in real time, the pollution level registered at a certain location, and allows transferring the gathered data to the Cloud server. In particular, it uses the Bluetooth interface to receive data from the Raspberry Pi, and the Wifi or cellular network interface to transfer data to the Cloud server.

The Cloud server has a web system that handles the information received from the Android device. The received data is stored in a database, which is then processed through the R Graph tool [19]. Finally, a detailed report is made available to the administrator of the web server.

IV. APPLICATIONS

EcoSensor is a platform composed by two integrated modules, (i) mobile elements responsible for capturing pollution values, and (ii) a cloud server responsible for storing and

processing the collected data. The mobile element is controlled via an Android-based application, and the server is managed by a Cloud Application. Below we provide a detailed overview of both applications.

A. Android application

The Android application was developed using the Android Studio IDE. This application allows starting or stopping a trace, viewing captured data in real-time, uploading data to the server, and performing other management tasks.

Internally, the application has two parts: (i) a service that continually receives (via Bluetooth) the data sent by the sensor, and that stores it in an internal database; and (ii) a user interface that allows starting or stopping a trace data capture from the sensor. It also provides real-time feedback about pollution levels at the current location according to the AQI index [3]. 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 POST message using the JSON format. The application is shown in Figure 2.

B. Server application

The cloud server provides a web interface based on Wordpress, which allows the administrator to have full access to the information gathered in terms of trace handling, processing, and visualization.

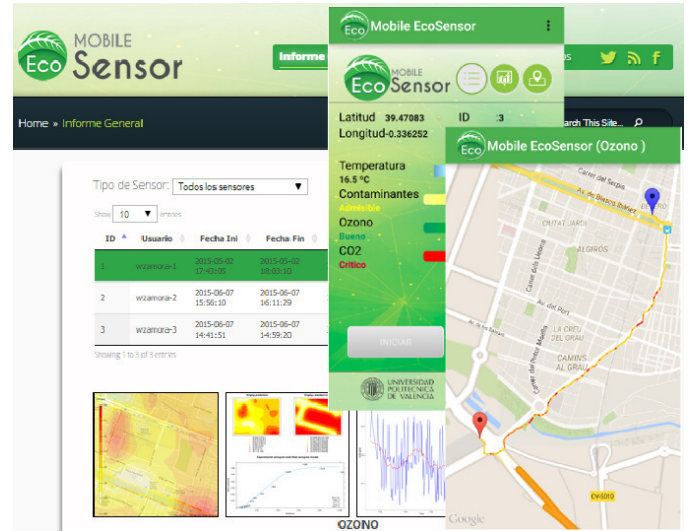


Fig. 2. EcoSensor: Cloud and mobile applications.

Once logged, the administrator views all uploaded traces, being able to perform different statistical analyses on the different data sets (CO2, Ozone, Air Pollution, and Temperature). For statistical analysis and report generation we relied on the R Graph tool, which offers us a way to generate graphics concerning:

- Pollution level.- It shows detailed pollution levels for a specific area.
- Kriging analysis.- It presents the kriging output: (i) interpolation result, (ii) kriging error, and (iii) variogram analysis.
- Captured data filtering.- it presents the adjustment/calibration process, showing the relationship between the original data and the resulting data.
- Data variation.- it provides a boxplot showing the overall data distribution.

The site is available at <http://www.ecosensor.net>, and its design is shown in Figure 2.

V. FINDING THE OPTIMAL MEASUREMENT STRATEGY

Once the proposed architecture has been fully developed and functionally tested, our next goal consists on determining the optimal strategy to collect air pollution data via mobile sensors.

With this purpose, we first analyzed the impact of mobility on sensor readings by comparing static measurements against mobile ones. In addition, we also 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 mobility. 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.

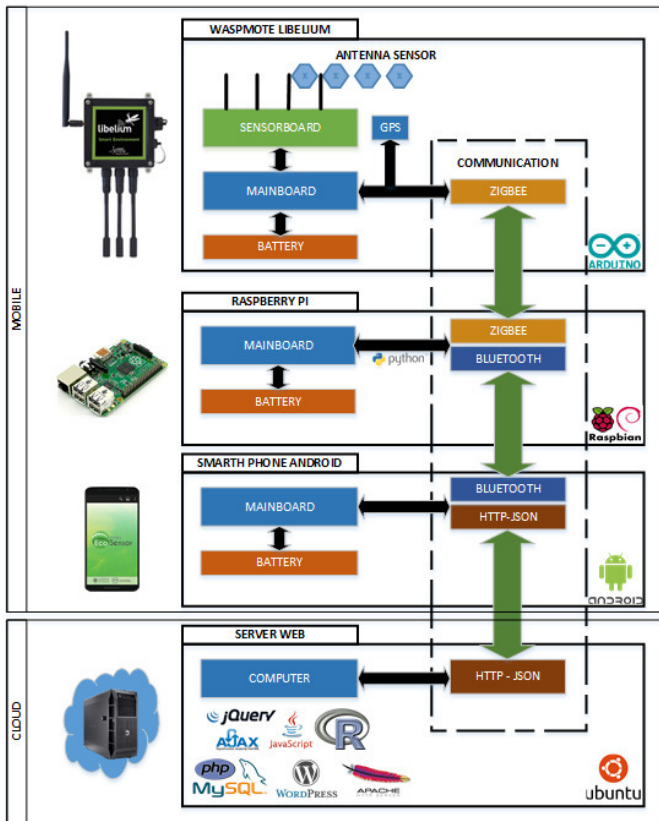


Fig. 1. EcoSensor Architecture.

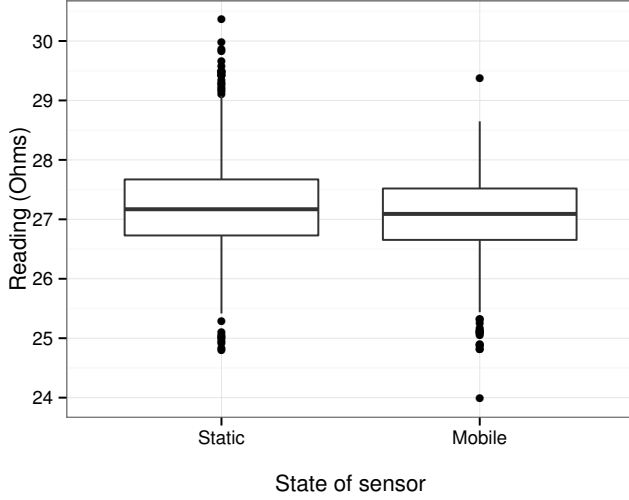


Fig. 3. Analysis of the variability of mobile sensor reading.

A. Optimal sensor positioning

To analyze the impact of mobility on the data capture process we performed different 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.

Figure 3 shows that mobility, at least at the speed used for testing, does not have a significant impact on sensor measurements. In particular, we find that values for the median and the inter-quartile range are quite similar to the ones obtained when the sensor is static.

B. 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 prediction error when generating the pollution map. For this purpose, we monitored the Technical University of Valencia campus (1.5km \times 0.6km) with a mobile sensor installed on a bike. The target pollutant used for the tests was ozone, due to its well known negative impact on health.

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 inter-sample 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 related datasets.

We used equation 1 to calculate the relative error using the initial trace (5s sampling) as a reference.

TABLE I
STATISTICAL SUMMARY OF THE TIME SAMPLING ANALYSIS.

Period	Mean	Std. Dev.	Error (ϵ_i)
5 sec.	60.31251	1.140371	-
20 sec.	60.37815	1.131514	0.04211991
40 sec.	60.36123	1.131782	0.07745983
80 sec.	60.45629	1.126616	0.08186843

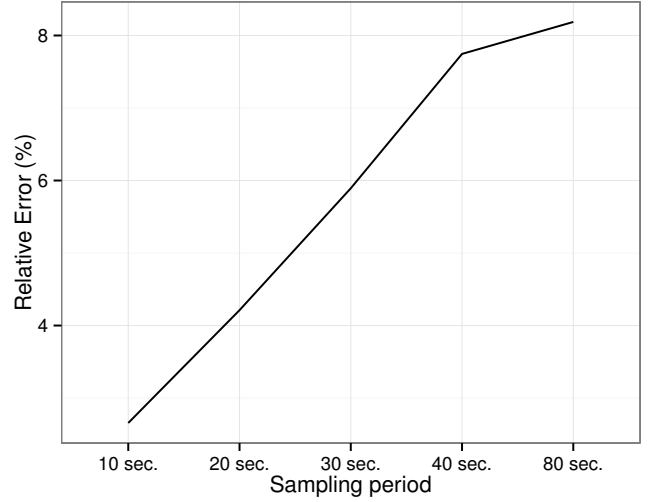


Fig. 4. Analysis of the relative error with respect to 5 sec. in different sampling period.

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

In this equation, ϵ_i represents the relative error 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 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 I, we can see that the mean and the standard deviation values are nearly the same in all cases, although the error increases when the inter-sample period increases.

Figure 4 shows how the relative error increases when the sampling period increases. Notice that, despite the distribution of values is similar, the mean error experiences an almost linear increase. Nevertheless, the relative error values are still relatively low since the kriging interpolation process also acts as an error filter, helping to approximate the mean value when lacking enough reference values.

Detailed maps for some traces (5 seconds, 20 seconds, 40 seconds, and 80 seconds) are shown in Figure 6, along with the data source used for kriging interpolation.

TABLE II
STATISTICAL SUMMARY OF THE SPATIAL SAMPLING ANALYSIS.

Dataset size	Mean	Std. Dev.	Error (ϵ_i)
100 %	60.31251	1.140371	-
72 %	60.49253	1.000335	0.06640983
50 %	60.66518	1.137273	0.11800896
42 %	60.51269	1.082692	0.14697632

By taking a look at the heatmaps, built using kriging interpolation techniques, we can clearly see that the level of detail experiences a degradation as we increase the sampling period. In particular we find that, although the pollution maps for 5 seconds and 20 seconds are quite similar, significant differences are observed when increasing the sampling period to 40 seconds, where the estimated ozone distribution 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 actual spatial distribution of those values.

C. Impact of spatial sampling on geostatistical predictions

We now evaluate the impact that spatial sampling will have on the predicted pollution map. In particular, we want to determine to which degree taking a shorter, less exhaustive path throughout the target area (i.e., 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 a shorter and yet valid trip, maintaining both start and end locations. As a result, we obtained traces with 72 %, 50 % and 42 % of the data. Similarly to the previous section we perform, for each dataset, a statistical analysis of the resulting data sets, also obtaining the pollution heatmap generated through kriging interpolation, and calculating the relative prediction error using equation 1.

Table II presents the statistical analysis. It shows the mean, the standard deviation, and the relative error, being the latter calculated using the initial dataset as reference.

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

Figure 5 shows how the relative error increases when spatial sampling decreases. Compared to the time sampling results in Figure 4, we find that now the error increase is more pronounced, meaning that reducing the route taken along the target area is prone to eliminate relevant samples, resulting in a less detailed pollution map.

Figure 7 shows detailed maps for datasets representing 100%, 72%, 50% and 42% of the data. Based on these heatmaps, we can clearly observe how spatial subsampling

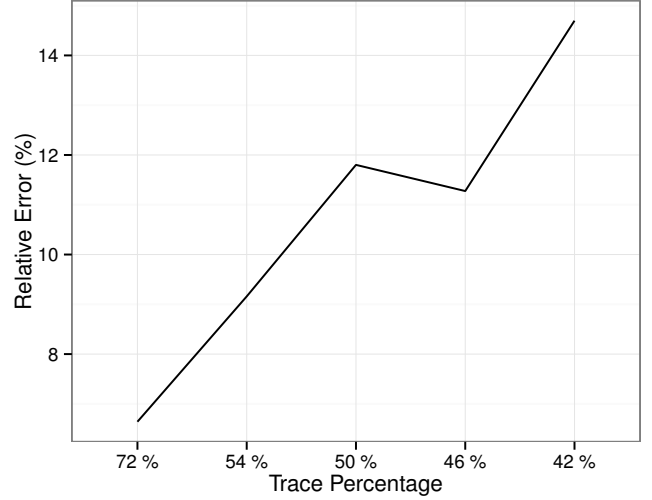


Fig. 5. Analysis of the relative error with respect to 100 % of trace

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, being time sampling granularity less, but yet somehow important. Finally, sensor orientation is the factor having the least impact on the obtained results.

VI. CONCLUSIONS

Nowadays, environment pollution monitoring has become a fundamental requirement for cities worldwide.

In this paper we propose EcoSensor, a complete solution for environmental monitoring that combines low-end sensors, smartphones, and cloud services to measure pollution levels with a high spatial granularity. Our approach is based on using a mobile sensor to provide pollution measurements, a smartphone to provide real-time feedback about air quality conditions, and also acting as a gateway to upload gathered data to the cloud server. EcoSensor additionally provides a cloud server for data processing and visualization.

To address the challenges associated with taking mobile measurements in a target area, we analyzed the influence of sensor orientation, as well as the impact that time and spatial sampling have on the sensing process. In particular, we varied the sampling period from 5 seconds to 80 seconds, and the overall path length from 100% to 42%, thereby being able to determine the most effective monitoring strategy. Experimental results show that the sensor orientation and the sampling period, within certain bounds, have very little 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.

The next steps in this research include improving the spatial interpolation process and developing a smaller mobile pollution sensor.

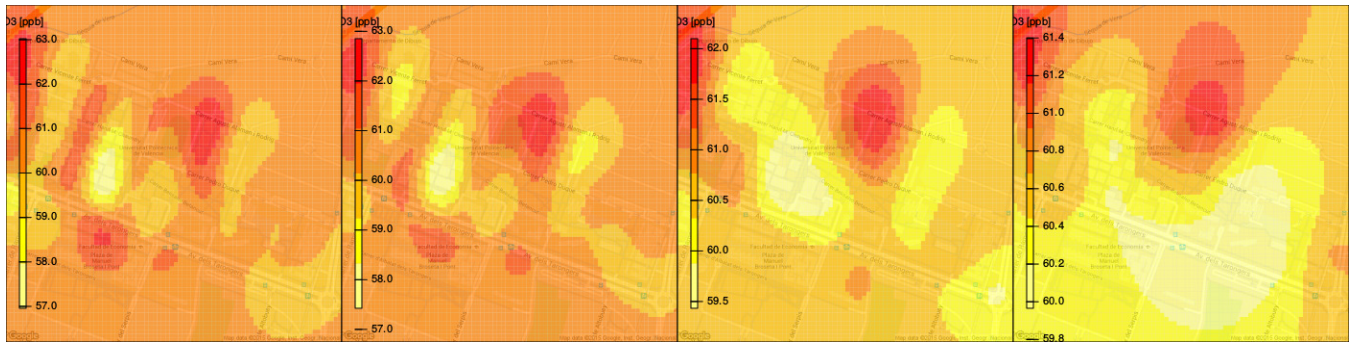


Fig. 6. Analysis of the ozone distribution for different sampling periods (5, 20, 40 and 80 seconds).

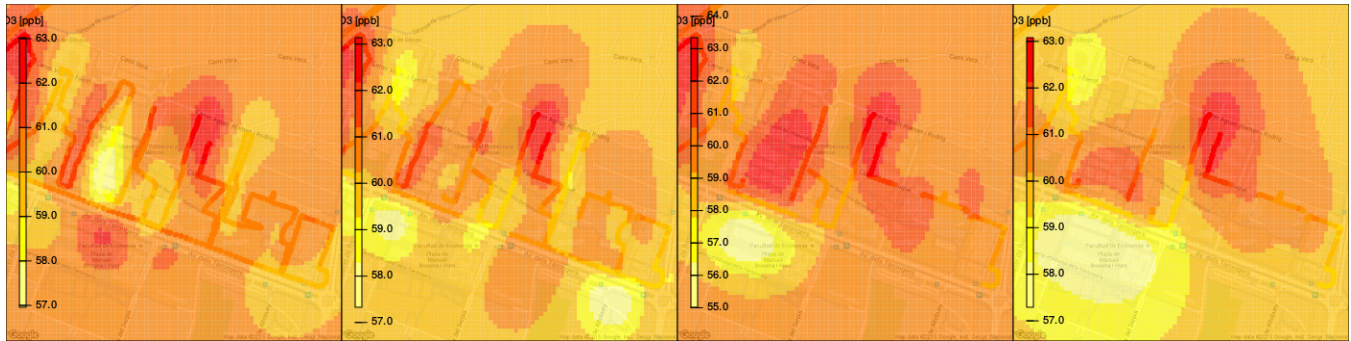


Fig. 7. Analysis of the ozone distribution with different trace percentages used (100 %, 72 %, 50 % and 42 %)

ACKNOWLEDGMENT

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", "Universidad de Cuenca", "Universidad Laica Eloy Alfaro de Manabí - ULEAM" and "Programa de becas SENESCYT de la República del Ecuador".

REFERENCES

- [1] I. Date, "United States Environmental Protection Agency," 2008.
- [2] ETC/ACM, "European Topic Centre on Air Pollution and Climate Change Mitigation Available: <http://acm.eionet.europa.eu/>," 2015.
- [3] U. S. E. P. Agency, "Air Quality Index Available: <http://cfpub.epa.gov/airnow/index.cfm?action=aqibasics.aqi>," 2015.
- [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. 244–8, 2007. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/17435419>
- [5] T. M. Chen, J. Gokhale, S. Shofer, and W. G. Kuschner, "Outdoor air pollution: Particulate matter health effects," *American Journal of the Medical Sciences*, vol. 333, no. 4, pp. 235–243, 2007. [Online]. Available: <http://go.isi.net/000246509200008>
- [6] R. D. Brook, J. R. Brook, B. Urrutia, 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. 970, no. January 2013, pp. 1–19, 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 *Congress Proceedings. 48th ISOCARP-International Society of City and Regional Planners World Congress: Fast Forward: Planning in a (hyper) dynamic urban context*, 2012. [Online]. Available: http://www.isocarp.net/Data/case_studies/2215.pdf
- [10] H. B. Deng and L. Zhang, "Design on ZigBee Wireless Sensor Network Node," *Key Engineering Materials*, vol. 474–476, pp. 283–286, 2011. [Online]. Available: <http://www.scientific.net/KEM.474-476.283>
- [11] C. T. Calafate and B. Ducourthial, "On the use of mobile sensors for estimating city-wide pollution levels," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2015 International*. IEEE, 2015, pp. 262–267.
- [12] Y. Cheng, X. Li, Z. Li, S. Jiang, Y. Li, J. Jia, and X. Jiang, "AirCloud: a cloud-based air-quality monitoring system for everyone," in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems - SenSys '14*. ACM, 2014, pp. 251–265. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2668332.2668346>
- [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*. ACM, 2013, pp. 1436–1444. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2487575.2488188>
- [14] Libelium, "Waspote Plug and Sense Smart Environment Available: <http://www.libelium.com/es/products/plug-sense/>," 2015.
- [15] Arduino, "Arduino Available: <https://www.arduino.cc/>," 2015.
- [16] Raspberry-Pi, "Raspberry Pi Available: <https://www.raspberrypi.org/>," 2015.
- [17] Zigbee, "Arduino Available: <http://www.zigbee.org/>," 2015.
- [18] Bluetooth, "Arduino Available: <http://www.bluetooth.com/>," 2015.
- [19] R-Foundation, "R Project Available: <https://www.r-project.org/>," 2015.