ATTI DEI DIRETTORI DELLE SEZIONI DI RICERCA SCIENTIFICA O TECNOLOGICA, DELLE STRUTTURE TECNICHE DI SERVIZIO E DEI DIRIGENTI

Atto n. GEO 460 ADW del 29/06/2023

Oggetto: **Servizio di pubblicazione, in open access, dell'articolo scientifico dal titolo: "Design and Implementation of a crowdsensing based air quality monitoring open and FAIR data infrastructure" sulla rivista Processes. Importo complessivo € 2.040,13, IVA esclusa. Oneri di sicurezza interferenziale non soggetti a ribasso d'asta: € 0,00. Affidamento diretto, ai sensi dell'art. 1, comma 2, lettera a) della L. n. 120/2020, così come modificata dalla L. n. 108/2021 e s.m.i., alla società MDPI, di Basel (CH), VAT NO. CHE-115.694.943, CIG n. ZD63BBBDED.**

IL DIRETTORE DELLA SEZIONE DI RICERCA SCIENTIFICA GEOFISICA

Richiamata la deliberazione CdA n. 86 dd. 08/10/2020 "Procedura comparativa per l'individuazione dei Direttori delle Sezioni e/o Centri di Ricerca Scientifica e/o tecnologica: Sezione di Ricerca Scientifica "Geofisica" – GEO: nomina del candidato vincitore" con la quale si è disposto la nomina a Direttore della Sezione di Ricerca Scientifica "Geofisica" – GEO, del dr. Fausto Ferraccioli, con decorrenza 01/11/2020 e termine al 31/10/2024;

vista la determinazione della Direttrice Generale n. 199 ADW del 12/05/2022 avente ad oggetto: *"Delega temporanea in materia di contratti pubblici di lavori, servizi e forniture ai Direttori delle Sezioni di ricerca/Centri di ricerca e delle Strutture Tecniche di servizio e ai Dirigenti Amministrativi dell'OGS";*

esaminata la richiesta di fornitura n. GEO 88/2023, completa di versione draft, predisposta dal dott. Poalo Diviacco, della Sezione GEO (Allegato n. 1), per la pubblicazione dell'articolo scientifico dal titolo: "Design and Implementation of a crowdsensing based air quality monitoring open and FAIR data infrastructure", sulla rivista Processes, autori Paolo Diviacco, Massimiliano Iurcev, Rodrigo Carbajales, Alberto Viola e Nikolas Potleca;

tenuto conto che la scelta della rivista è legata al prestigio dell'editore e al ranking della rivista;

ricevuta e ritenuta congrua, la fattura n. 2411337 di data 20/06/2023, inviata dalla società MDPI, di Basel (CH), contenente il corrispettivo per il servizio richiesto, al costo complessivo di € 2.040,13, IVA esclusa (Allegato n. 2);

considerato l'art. 1 comma 2, lettera a) della L. n. 120/2020, così come modificato dalla L. n. 108/2021 e s.m.i., laddove si dispone che le stazioni appaltanti possano procedere, per importi inferiori a € 139.000,00, ad affidamenti diretti anche senza previa consultazione di due o più operatori economici;

verificato inoltre che risultano applicabili le disposizioni di cui all'art. 10, comma 3, del D.Lgs 25/11/2016, n. 218 (Semplificazione delle attività degli Enti Pubblici di ricerca ai sensi dell'art. 13 della L. 7/8/2015, n. 124), in quanto il servizio in oggetto è funzionalmente collegato e destinato ad attività di ricerca nell'ambito del progetto sopra indicato, non sussistendo quindi l'obbligo di ricorso al mercato elettronico della Pubblica Amministrazione (MePA);

tenuto conto inoltre, delle disposizioni contenute nella c.d. Legge di Stabilità 2016 (L. 28/12/2015, n. 208), modificata con la c.d. Legge di Stabilità 2019, laddove è possibile effettuare acquisti di beni e servizi di valore inferiore ad € 5.000,00 (IVA esclusa) al di fuori del MePA Consip;

ritenuto inoltre applicabile nella fattispecie quanto previsto dal punto 3.7 delle linee guida ANAC n.4, laddove è ammessa la deroga al principio di rotazione per importi inferiori a € 1.000,00, qualora l'affidatario, nel precedente appalto, abbia dimostrato piena affidabilità, comportamento virtuoso nell'esecuzione dello stesso, ed abbia praticato prezzi in linea con quelli del settore di mercato di riferimento, circostanze presenti nella procedura di cui trattasi;

richiamato inoltre il punto 5 del parere del Consiglio di Stato n. 01312/2019 dd. 30/04/2019, con il quale si condivide l'innalzamento, da € 1.000,00 ad € 5.000,00, della soglia entro la quale è possibile, con scelta motivata dalle circostanze di cui sopra, derogare al principio di rotazione;

ritenuto pertanto di procedere all'affidamento del servizio per quanto in oggetto alla società MDPI, di Basel (CH);

considerato l'esiguo importo dell'affidamento, la richiesta di prestazione della garanzia sarebbe antieconomica rispetto al valore del contratto;

ritenuto pertanto, date le condizioni di cui sopra, di non richiedere la garanzia definitiva nei confronti dell'operatore economico ai sensi dell'art. 103 del D.Lgs. n. 50/2016 e s.m.i.;

considerato che la spesa pari a stimati € 2.040,13, IVA esclusa, per il servizio di cui trattasi, trova idonea copertura, sul capitolo 52002, art. 64 commessa 6750 (ION), del Bilancio di previsione 2023, di pertinenza della Sezione GEO;

richiamato altresì l'art. 31 del D.Lgs. n. 50/2016 e s.m.i., e considerato che lo scrivente svolge le funzioni di Responsabile Unico del Procedimento mentre la sig.ra Simona Cassaro funge da Responsabile dell'Istruttoria;

preso atto che non ricorrono nel caso di specie motivi di incompatibilità o di conflitto di interesse, nei confronti dello scrivente, ai sensi della vigente normativa in materia di contrasto alla corruzione;

accertato che, in conformità a quanto disposto dall'AVCP con determina n. 3/2008, il servizio di cui trattasi non comporta rischi da interferenza, per cui non è stato redato il DUVRI;

visti gli articoli 32, commi 6 e 7, e 33, del D.Lgs. n. 50/2016 e s.m.i.;

visto il Regolamento di amministrazione, finanza e contabilità dell'OGS;

tutto ciò premesso e considerato;

D E T E R M I N A

1. di procedere, per i motivi esposti in premessa, all'affidamento diretto del servizio di pubblicazione, in open access, dell'articolo scientifico dal titolo: "Design and Implementation of a crowdsensing based air quality monitoring open and FAIR data infrastructure" sulla rivista Processes, alla società MDPI, di Basel (CH), VAT NO. CHE-115.694.943, ai sensi dell'art. 1, comma 2, lettera a) della L. n.

120/2020, così come modificata dalla L. n. 108/2021 e s.m.i., CIG n. ZD63BBBDED, al prezzo complessivo di € 2.040,13 IVA esclusa;

2. di stipulare il contratto nelle more dell'effettuazione di successivi eventuali ulteriori controlli a campione delle verifiche della veridicità delle dichiarazioni rilasciate dall'operatore economico, inserendo nelle clausole di risoluzione il pagamento all'impresa del corrispettivo pattuito solo con riferimento alle prestazioni già eseguite e nei limiti dell'utilità ricevuta e di applicazione di una penale in misura pari al 10% (dieci per cento) del valore del contratto, qualora l'esito delle suddette verifiche risulti negativo;
3. per le ragioni sopra descritte, di non richiedere all'affidatario la prestazione della garanzia definitiva di cui all'art. 103 del D.Lgs. n. 50/2016 e s.m.i.;
4. di dare atto che la spesa complessiva di € 2.488,96, IVA inclusa, per il servizio di cui trattasi, trova idonea copertura, sul capitolo 52002, art. 64 commessa 6750 (ION), del Bilancio di previsione 2023, di pertinenza della Sezione GEO;
5. che il Responsabile del Procedimento in oggetto è il sottoscritto mentre la sig.ra Simona Cassaro funge da Responsabile dell'Istruttoria;
6. di dare atto che il presente procedimento sarà soggetto ad avviso di post-informazione mediante pubblicazione sul profilo istituzionale del committente, nella sezione Gare e appalti;
7. di impegnare la spesa complessiva di € 2.488,96 sui capitoli di seguito elencati:

Eser.	EPF	Cap.	Art.	CIG	Cod Bilancio	Importo	Soggetto
2023	2023	52002	64	ZD63BBBDED	52002	2.488,96	MDPI AG cod.fisc. / p.i. CHE115694943

IL DIRETTORE
Fausto Ferraccioli

Elenco firmatari

ATTO SOTTOSCRITTO DIGITALMENTE AI SENSI DEL D.P.R. 445/2000 E DEL D.LGS. 82/2005 E SUCCESSIVE MODIFICHE E INTEGRAZIONI

Questo documento è stato firmato da:

NOME: FERRACCIOLI FAUSTO
CODICE FISCALE: FRRFST70A29D969K
DATA FIRMA: 29/06/2023 13:29:47
IMPRONTA: 7A07B48892453EFFF1B581519A219D9C2920672CED467A60A7E5ADB91EA51D47
2920672CED467A60A7E5ADB91EA51D476A012892F28DE7EFD46EA9EDF0631FB2
6A012892F28DE7EFD46EA9EDF0631FB2C0C001989398871713AA28F7F99D03E7
C0C001989398871713AA28F7F99D03E793F7654C6025889191D6A14334859557

MODULO RICHIESTA FORNITURA/SERVIZIO

LE RICHIESTE SARANNO NUMERATE A CURA DELL' UPA DELLA DFP O DAL GS DELLA SEZIONE

STRUTTURA ARTICOLAZIONE Infr. Gestione. Dati Geof. DATA RICHIESTA N. GEO 88/2023

RISERVATO ALL'UPA O AL G.S.

N° PROG.	INV.	CO NS.	DESCRIZIONE	COMMESSA	UNITA' DI MISURA	Q.TA'	IMPORTO UNITARIO	AL. IVA	PREZZO	CONSIP NO	CONSIP Esaurito	CONSIP Prezzo
			APC per pubblicazione su MDPI Processes Design and Implementation of a crowdsensing based air qual- 4 ity monitoring open and FAIR data infrastructure Paolo Diviacco *, Massimiliano Iurcev, Rodrigo José Carbajales, Nikolas Potleca, Alberto Viola	6750 ION		1			2040,13			

NOTE _____

IL RICHIEDENTE

RESP. STR./ARTICOLAZIONE

RESP. COMMESSA

CIG N.

VOCE DI SPESA AMMESSA A RENDICONTAZIONE ☐ Sì ☐ NO

DITTE INTERPELLATE

1

2

3

4

5

TOTALE IVA COMPRESA € € 2488,96

DISCIPLINARE PER L'ACQUISTO D.LGS. 50/2016

☐ Art. 36, II comma, lettera **a)**

☐ Art. 36, II comma, lettera **b)**

Altro

RICHIESTA DI RIMBORSO MEDIANTE SERVIZIO INTERNO DI CASSA

☐ spese per piccole riparazioni, manutenzioni di mobili e locali ☐ minute spese d'ufficio
☐ spese di vettura ☐ spese postali ☐ spese per giornali, pubblicazioni, periodici e simili

FIRMA PER AVVENUTO RIMBORSO

FIRMA PER ACQUISTO TRAMITE SERVIZIO INTERNO DI CASSA

☐ IL DIRETTORE GENERALE ☐ IL DIR. DI SEZ. O STR. TECN. DI SERV.

☐ IL DIRIGENTE AMM.VO

CAPITOLO / ART.	N° IMPEGNO	FIRMA	DATA
52002 / 64			

AUTORIZZAZIONE ALL'ACQUISTO

NOTE RELATIVE ALL'AUTORIZZAZIONE DATA

☐ IL DIRETTORE GENERALE
☐ IL DIR. SEZ. O DI STR. TECN. SERV.
☐ IL DIRIGENTE AMM.VO

DATA **REGISTRAZIONE UFFICIO RAGIONERIA**

ORDINE EVASO ☐

Oggetto: Relazione RDF pubblicazione open access su MDPI Processes

Nell'ambito delle attivita' di citizen science dell'Infrastruttura di gestione dei dati geofisici della sezione GEO e' stato sviluppato un sistema di monitoraggio della qualita' dell'aria nella citta' di Trieste che copre tutti gli aspetti, dalla misurazione alla trasmissione all'elaborazione alla mappatura geografica dei dati su portale web. Il sistema e'stato reso possibile grazie ad una stretta collaborazione con la Trieste Trasporti. Nel dettaglio di questo articolo abbiamo focalizzato il problemi e le implementazioni tecnologiche del sistema. E' stato evidenziato come con il sistema proposto sia possibile mappare con grande precisione sia geograficamente che temporalmente l'evoluzione della distribuzione del particolato. Nel dettaglio e' stato descritto il sistema di acquisizione ed i metodi di calcolo e rappresentazione web dei dati.
L'esperienza e' stata descritta in questa pubblicazione che ha avuto reviews molto positivi. Dopo delle minor revisions l'articolo e' stato accetto. La rivista ha un Impact factor di circa 3.3. Ed impone essendo open access il pagamento dei costi di produzione pari a 2040,13 Euro

La spesa gravera' sulla commessa 6750 ION.

Quantità	Bene
1	APC per pubblicazione su MDPI Processes Design and Implementation of a crowdsensing based air quality monitoring open and FAIR data infrastructure **Paolo Diviacco *, Massimiliano Iurcev, Rodrigo José Carbajales, Nikolas Potleca, Alberto Viola, Mihai Burca and Alessandro Busato**

In fede

Paolo Diviacco
Coordinatore infrastruttura Distribuzione dati

processes

Type of the Paper (Article)

Design and Implementation of a crowdsensing based air quality monitoring open and FAIR data infrastructure

Paolo Diviacco [1,*], Massimiliano Iurcev [1], Rodrigo Carbajales [1], Alberto Viola [1] and Nikolas Potleca [1]

[1] National Institute of Oceanography and Applied Geophysics, Borgo Grotta Gigante 42/C, 34010 Sgonico, Italy
* Correspondence: pdiviacco@ogs.it; Tel.: +39-040-2140380

Abstract: This work reports on the development of a real-time vehicle sensor network (VSN) system and infrastructure devised to monitor particulate matter (PM) in urban areas within a participatory paradigm. The approach is based on the use of multiple vehicles where sensors, acquisition and transmission devices are installed. PM values are measured and transmitted using standard mobile phone networks. Given the large number of acquisition platforms needed in crowdsensing, sensors need to be low-cost (LCS). This sets limitations in precision and accuracy of measurements that can be mitigated using statistical methods on redundant data. Once data are received, they are automatically quality controlled, processed and mapped geographically to produce easy to understand visualizations that are made available in almost real-time through a dedicated web portal. There, end users can access current and historic data and data products. The system has been operational since 2021 and has collected over 50 billion measurements, highlighting several hotspots and trends of air pollutions in the city of Trieste (North-East Italy). The study concludes that (i) this perspective allows to drastically reduce costs and considerably improve the coverage of measurements; (ii) for an urban area of approximately 100.000 square meters and 200.000 inhabitants a large quantity of measurements can be obtained with a relatively low number (5) of public buses; (iii) that a small number of private cars, although less easy to organize, can be very important to provide infills in areas where buses are not available; (iv) appropriate corrections for LCS limitations in accuracy can be calculated and applied using reference measurements taken with high quality standardized devices and methods; (v) analyzing the dispersion of measurements in the designated area it is possible to highlight trends of air pollution and possibly associate them with traffic directions. Crowdsensing and open access to air quality data can provide very useful data to the scientific community but also have great potential in fostering environmental awareness and the adoption of correct practices by the general public.

Keywords: Particulate matter; crowdsensing; citizen science; low-cost sensors; infrastructure.

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

1. Introduction

The World Health Organization defines air pollution as the "contamination of the indoor or outdoor environment by any chemical, physical or biological agent that modifies the natural characteristics of the atmosphere." [1]. Epidemiological evidence suggests that polluted air is one of the leading factors associated with development of respiratory illness, cardiovascular disease, and lung cancer [2]. At the same time, air pollution directly and indirectly affects the climate and damages buildings and cultural heritage [3], [4]. Many countries have introduced specific legislation setting strict objectives for air quality. In the United States, this was implemented in the 1970s with the Air Quality Act [5] and later in Europe, with the 2008/50/EC directive [6]. Notwithstanding the fact that in general air quality has improved a lot since then, there are still several hot spots of air pollution in most of the western countries [7], [8]. Recently, several analysts highlighted the risks that, due to the current geopolitical situation and the shortage of natural gas, the resulting increase in the use of solid fuels could worsen the situation [9]. Indeed, while the combustion of natural gas contributes to the formation of smog and acid rain, particulate matter (PM) emissions are generally low [10]. On the contrary, liquid and solid fuels combustion produces large quantities of PM and high concentrations of sulfur and heavy metals [11].

Air quality monitoring is generally performed by government agencies using standardized methods and devices at specific fixed locations in order to have reliable long time series that could be considered as reference measurements. The position of such monitoring stations is linked to the specific problem to be considered, be it traffic congestion in an urban area or an industrial site or any other issue that could be relevant to public health.

Reference methods are intrinsically expensive and need well trained personnel. As a consequence, there are limitations in the possible number of stations to install. In order to reconstruct the geographic distribution of the quality of air in an area, measurements at the sparse stations can be inter/extrapolated using statistical [12] or modeling techniques [13]. Although generally very accurate, these methods can be problematic where high gradients are present. In such cases, the possibility to increase the geographic and temporal coverage and resolution of phenomena could be very helpful.

In this perspective, a new paradigm can be introduced that has already been successfully applied in several scientific fields. Since the seminal work of Irwin [14], a large number of initiatives have, in fact, flourished that aim at enrolling resources from outside the scientific community and employ them within several research activities [15]–[19]. This new approach is generally referred to as citizen science, although slightly different definitions may be more suited to each specific application. In this work, we prefer to use the term 'crowdsensing', which refers to a technique where a large number of volunteers offer their help in acquiring and sharing measurements taken with devices or software provided by the project designer. Within this perspective, a large literature exist on the use of air quality sensors [20] for several air quality parameters [21][22][23] both indoor [24][25][26] and outdoor [27][28] . Within this work we will focus on mobile crowdsensing [29], where PM acquisition devices are installed on vehicles such as cars, vans or public buses. As a result, the availability of a large number of vectors has the potential to radically increase the amount of data available and improve geographic and temporal coverage and resolution.

On the other hand, participative research activities such as crowdsensing and citizen science, have the possibility to deliver benefits well beyond scientific outcomes. Starting from the importance of understanding how the general public is informed about environmental topics, such as climate change or air quality, it is easy to understand that if the discussions remain confined within the scientific community it is very unlikely that the general public will be able to take informed actions in order to mitigate those phenomena. Effective and correct communication via the mass media and the Internet is therefore necessary to spread correct messages, which, unfortunately, is often not the case because mass media and in particular the social media may in the least be partial if not altogether manipulative.

On the contrary, active participation in research activities by volunteers, together with the possibility to freely and easily access reliable data and information on environmental issues can have a wide range of positive effects. These can span from an increase of trust in the scientific community, to the improvement of awareness and engagement of citizens in environmental issues [30] up to their empowerment in steering political and economic decisions [31], [32]. While improvement in subject-matter knowledge and stronger scientific literacy is generally easy to be traced as a participative research outcome, the actual impact on policy making of such initiatives is not always easy to fully understand and is sometimes a matter of debate [33], [34].

To support citizen science and crowdsourcing activities in the field of environmental monitoring several, often intermingled, aspects have to be considered. Each of them can constitute on their own a topic for a considerable analysis. In previous works [35]–[39], we have considered some of these aspects and the scientific results that we were able to obtain exploiting them. Here, we describe the technological aspects of our work in the hope that our experience could prove beneficial to others intending to replicate or eventually improve what we have been able to build so far.

2. Materials and Methods

Within this work, we will describe a system and infrastructure we developed that, leveraging the crowdsensing paradigm, allows monitoring air quality and represent its geographic distribution on a web portal in real time. The initiative is named "COCAL" after the dialectal term used for seagulls in the city of Trieste (Italy), where it has been developed and first deployed. The reason for using a seabird name comes from the fact that Trieste is a coastal city where applications of the crowdsensing paradigm can be envisaged in multiple environments. As a matter of fact, the first trials of COCAL were focused on monitoring marine parameters such as temperature, pH, salinity and dissolved oxygen. Further information on such trials can be found in [35], [36]

Trieste is located at the Nort-East tip of the Adriatic Sea and occupies a NW-SE trending elongated area of approximately 90 square kilometers between the Sea and the Karst plateau which acts as a barrier to air masses. The city center lies in a restricted area characterized by economic activities linked to the tertiary sector and tourism, and that is separated from an industrial area located in a SE sector. The wind regime is characterized in winter by a strong NE wind called Bora, that can effectively move polluted masses from the city to the sea and other nearby regions[13][40]. In summer the sea breezes are the prevalent factor conditioning the behaviour and position of the polluted air masses[41].

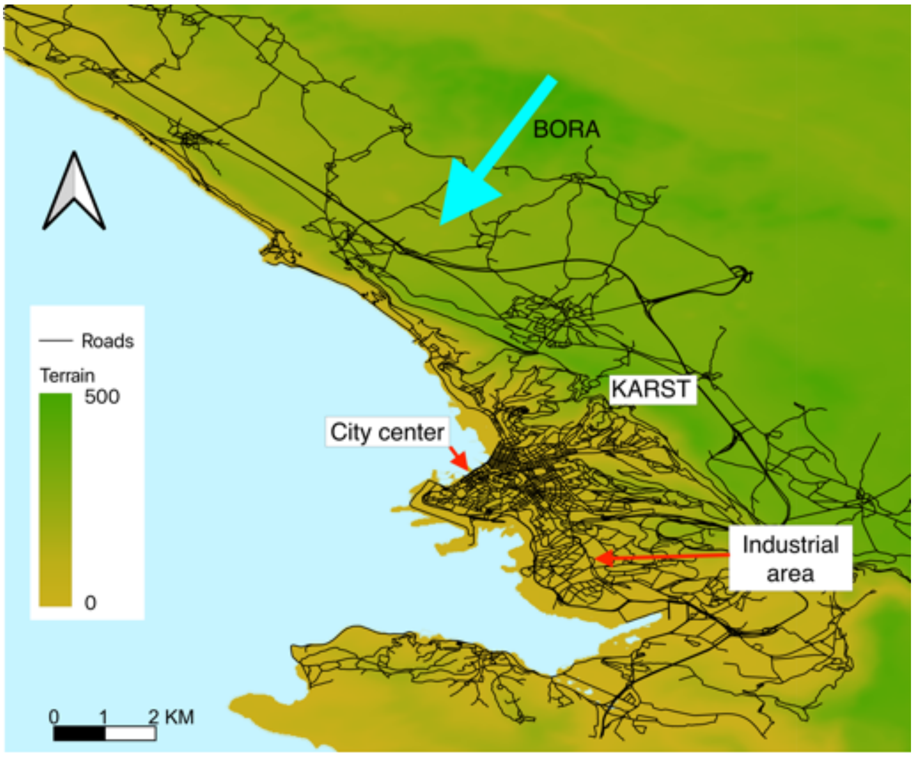


Figure 1. The area where the COCAL system is installed

Within this work we focus on monitoring PM only, but we are currently working on extending the method to other pollutants. It is worth highlighting that all the data acquired within this initiative and in other crowdsensing initiatives are gathered in an integrated database and managed within a fully FAIR compliant perspective following international standards as mandated by ISO and OGC.

2.1. PM Sensors

Given the importance, in crowdsensing, of using a large number of acquisition devices, it is evident that increasing their number will inevitably imply increasing the overall cost of the initiative. Under this approach, in fact, it is not possible to use conventional PM monitoring techniques such as filters and gravimetric mass detection, which are very expensive and based on standardized procedures that can only be performed by trained personnel. Therefore, low-cost sensors (LCS) are needed. New technologies have emerged that use laser scattering, which relates the waveform of the scattered light to the diameter and number of particles, enabling real-time and continuous measurements of particulate matter.

A detailed description of the technologies behind PM sensors is beyond the scope of this work and can be found in other works such as for example [42], [43]. Suffice it to say that these PM sensors consist of a fan, generally connected to a small tube, that pushes air into the sensing box. Light from a laser diode is scattered by the particles. This scattered light is received by a photodiode, which can estimate the concentration of each type of particle by classifying and counting the number of pulses detected.

Within COCAL, we use the SDS011 PM sensor from Nova Fitness Co. which enables simultaneously measuring both $PM_{2.5}$ and PM_{10} levels at a very low cost.

2.1.1. LCS performances

The major advantages of LCS in terms of price and portability come at the expense of limitations in precision and accuracy [43]–[45].

Many LCS manufacturers and models are available on the market, and detailed comparisons between them can be found, for example, in [12],[42],[24]. These works highlight that, in addition to the limitations in precision and accuracy, it is very important to consider the environmental conditions in which LCS sensors operate. For example, because these sensors do not have sample conditioning equipment, they are susceptible to drifts due to Relative Humidity (RH), which can affect the hygroscopic growth of particles and distort measurements [46][27]. Results of these studies demonstrate that among LCS, the issue of quality and precision of the specific brand and model of sensors can be as relevant as the intrinsic limitations of the technology employed, the issues related to the deployment in the designated environment, and the environmental conditions.

In this perspective, to monitor the environmental conditions in which the acquisition takes place, together with the PM LCS we also use a Dallas Semiconductor DS18B20 one-wire communication sensor with waterproof protection outside the acquisition box, while internally we use a Bosch BME280 temperature, pressure, and RH sensor connected to the board via the I2C bus.

In [42] useful references can be found to understand the performances of the SDS011 sensor and other similar sensors under controlled laboratory conditions. Results of that work confirm that the SDS011 sensor is suitable for use within the COCAL project since it performs reasonably well in comparison with similar or even more expensive sensors. At the same time, downsides have been identified such as a general trend to underestimate PM values and the presence of a delay in the timing of measurements.

2.1.2. Statistical analysis

Following [44], it is particularly important to understand the behavior of the LCS used in this work under real-word conditions. In order to devise possible mitigation strategies for the issues introduced by the use of LCS, following the protocols suggested by US EPA and [43] we designed two experiments, and namely: (i) a study of the behavior of LCS in a highly variable PM levels environment to evaluate precision and (ii) a colocation-based evaluation of LCS with a reference measurement station managed by the regional environmental agency ARPA-FVG, in order to evaluate accuracy. It should be noted that during these tests, data completeness of the COCAL system, meaning the ability to avoid gaps in measurements and data transmission, has always been very high, with almost negligible glitches and well above the 75% threshold recommended by US EPA.

2.1.2.1 LCS precision

To assess the precision of LCS, we studied the recordings of three co-located LCS in a highly variable PM concentration environment. Following US EPA standards and procedures mentioned by [43], precision can be estimated using The Standard Deviation (SD) and the Coefficient of Variation (CV). The SD shows a value of 2.15 for PM_{10} and a value of 1.17 for $PM_{2.5}$, while the CV shows a value of 24.70% for PM_{10} and 22.77% for $PM_{2.5}$.

US EPA recommends a target SD less or equal to 5 $\mu g/m^3$ and a CV less or equal to 30%. The tests we conducted therefore assess that LCSs used in COCAL match the recommendations for precision.

2.1.2.2. LCS accuracy

To test the accuracy of the selected sensors, we placed a COCAL box at a short distance from a certified reference air quality station (ARPA-FVG station 'Rosmini'). The reference PM values of this station are made available through an API on the official ARPA-FVG website as daily average values only. The measurements and comparison took place from mid-March 2022 to the end of April 2023.

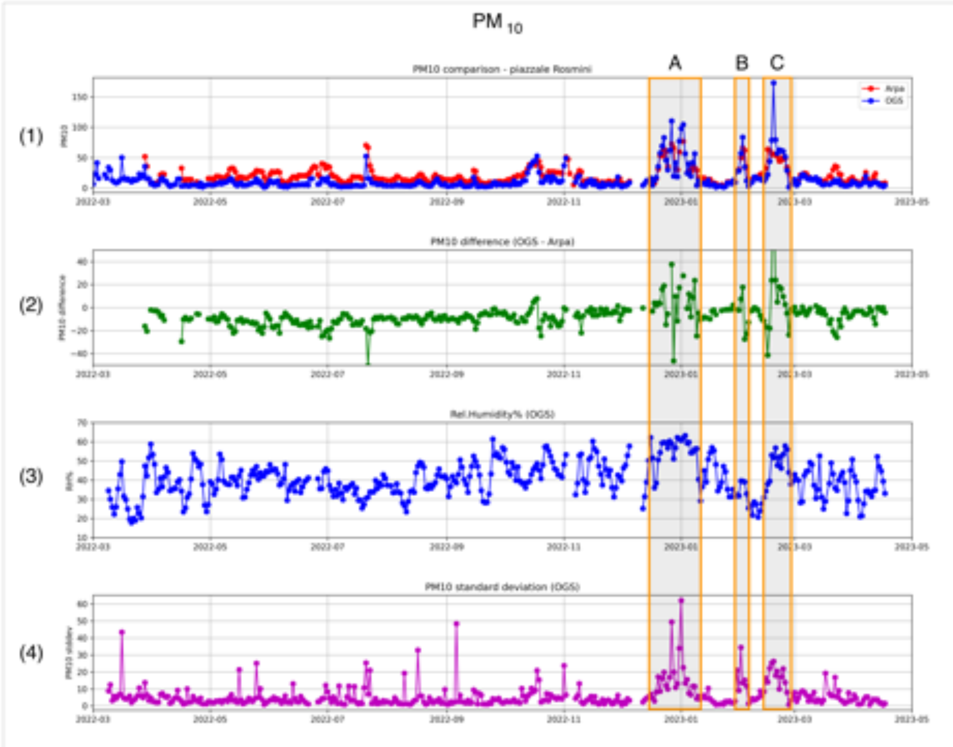
In Figure 2, one can compare measurements from the reference station and those taken during the same period with the COCAL box located close to the reference station.

Figure 2 is divided in two boxes. The upper box shows data and statistics for PM_{10}, while the lower one focuses on $PM_{2.5}$. In each box, the first graph (n.1) shows in red the time series of the reference daily average values as made available from the ARPA-FVG web site, while the daily average values of the COCAL box located close to the reference station are plotted in blue.

High PM values were measured at the end of July 2022. Unfortunately, these were not outliers, but the effects of a large forest fire that occurred for several days at a distance of about 20 km from the test site.

It is possible to note that COCAL measurements are generally lower than the reference measurements; however, during the first months of 2023, in three specific events (identified by boxes A, B and C) this behavior reverses. To understand better the performances of LCSs, Figure 2 also shows the difference between the reference and the COCAL time series (graph n.3), the RH time series (graph n.3) and the time series of the standard deviation, of all COCAL measurements acquired near the reference station, calculated on a daily basis.

It is interesting to note that during the A, B and C events, the standard deviation of COCAL measurements increases. In two of these cases (A and C), this can be understood as owing to the sensitivity of the LCS to RH. In fact, the time series of the RH in that periods exceeded 60% while elsewhere, when LCS underestimated the reference measurements, the RH remains below this threshold. In the cases of event B, however, where RH is low, a different explanation is needed. This could be found in the different technologies and rate of sampling of the LCS and the reference acquisition system. COCAL boxes acquire data every ten seconds, which means that rapid variations in the actual PM concentrations can effectively be captured, increasing at the same time the standard deviation of the set of daily measurements. Reference systems sample much more slowly so that, even if very reliable, they can overlook rapid phenomena so that in the comparison LCSs data appear to be dispersed.



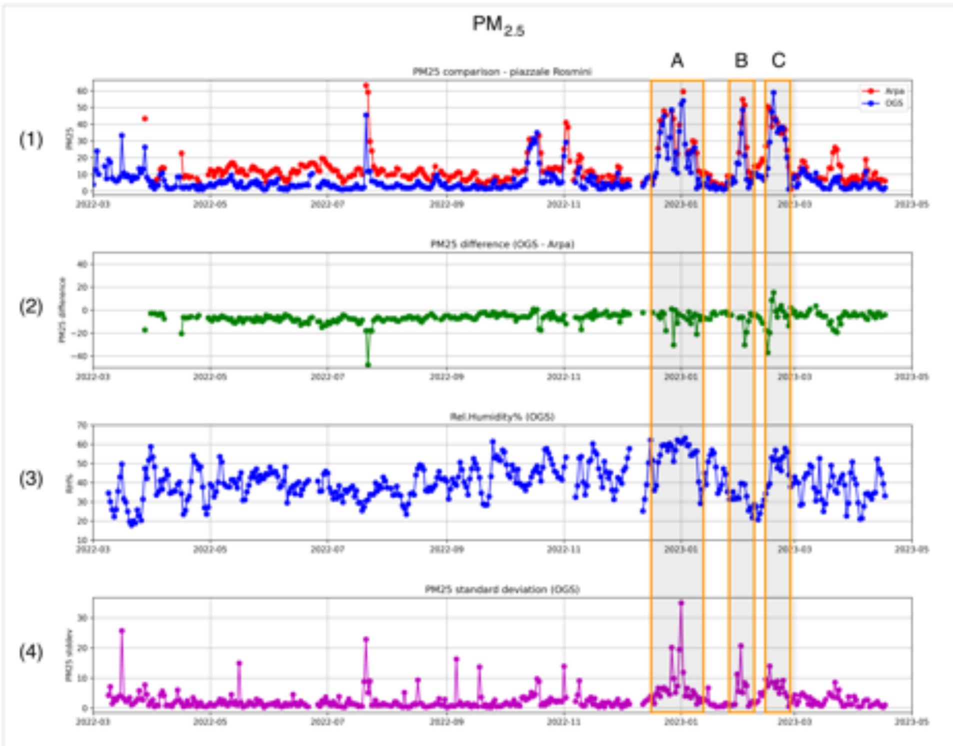


Figure 2. Comparison of 15 months of PM measurements from a certified reference station with measurements taken during the same period with a COCAL system located near the reference station. The upper set of graphs reports PM_{10} measurements, while the lower one the $PM_{2.5}$ measurements. In both cases, graph n.1 shows the daily average COCAL measurements in blue, while the reference daily value is plotted in red. Graph n. 2 shows the difference between the two time series.

Graph n. 3 shows the RH time series. Graph n. 4 shows the standard deviation of COCAL measurements calculated day by day. As can be seen, COCAL measurements are generally lower than the reference measurements. During the first months of 2023 in three specific events (identified in the graph by the boxes A, B and C), this behavior reverses. It is interesting to note that during these events the standard deviation of measurements also becomes high. In two of these cases (A and C), this can be understood as owing to the sensitivity of the LCS to RH, but in the case of event B the RH is low.

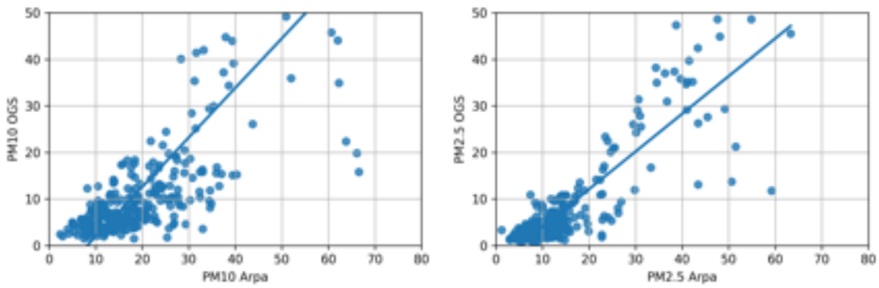


Figure 3. Scatter plot comparing Reference PM measurements and COCAL measurements

Following US EPA standards and procedures mentioned by [32], accuracy can be estimated using the Coefficient of Determination (R^2), Slope (m), Intercept (b), Root Mean Square Error (RMSE) and the Normalized Root Mean Square Error (NRMSE). Results for the mentioned tests for PM_{10} and $PM_{2.5}$ are shown in Table 1, while Figure 3 provides a snapshot of the comparison of the LCS and reference measurements.

Table 1 LCS accuracy metrics.

Parameter	**PM_{10}**	**$PM_{2.5}$**
Coefficient of Determination (R^2)	0.45	0.25
Slope	1.06	0.80
Intercept (b)	-8.73	-4.09
Root Mean Square Error (RMSE)	13.76	8.73
Normalized Root Mean Square Error (NRMSE)	66.04	60.37

Considering US EPA recommendations, these results can be problematic. In fact, considering R^2 this parameter is recommended to exceed 0.70 while the analysis reveals lower values. The target slope should be approximately 1 +- 0.35, a condition that is instead respected by the LCSs. Similarly, the Intercept parameter performs relatively well for $PM_{2.5}$ sensors, while PM_{10} sensors do not fall within the recommended range since EPA recommends a value between -5 and +5. Following EPA standards, RMSE should be lower than 7 µg/m³, and again, also here $PM_{2.5}$ scores rather well, whereas we measured twice the threshold for PM_{10}. In addition, unfortunately also the NRMSE results are too high, being around 60%, while EPA recommends a value less than 30%.

It can therefore be said, that accuracy-wise, the LCSs perform rather poorly, both for PM_{10} and $PM_{2.5}$ and that a correction mechanism is needed to obtain results that could be comparable with the official ones. At the same time, it is possible to say that, since the precision is reasonably good, a geographic distribution of LCS measurements built on the integration of multiple COCAL boxes should reasonably be capable of highlighting general trends and pinpointing local anomalies.

2.1.3 LCS performances improvement

The results of the tests are consistent with a wide literature on LCS performances. Several authors [47]–[49] highlighted the impact of the environmental conditions and in particular of the RH in the deviation between reference measurements and LCS. To address such problems, reference stations are equipped with a device that on heating air samples induces the water vapor condensed onto the particle to evaporate. This, of course, is not possible with LCSs. To compensate for this effect, several RH correction approaches exist such as for example: κ-Köhler theory derived factors, various types of regressions or machine learning methods. After an extensive survey of the existing literature, [44] maintains that such corrections are applied very seldomly, being a simple linear regression, in that case, the most used method. The same authors underline difficulties in accurately defining local parameters and accumulating knowledge from different cases and areas. It is also worth noting that RH itself can be a problematic parameter to measure, and that since COCAL is a VSN system, RH measurements taken with it can have further limitations.

Taking these considerations into account, and since the tests done in section 2.1.2.1 and 2.1.2.2 with the LCSs we use in this work, revealed that they perform reasonably well in terms of precision but unfortunately not well enough in terms of accuracy, we devised a specific and pragmatic two-step mitigation strategy to improve their performances.

The first step consists in filtering all measurements made in problematic conditions, for example, when RH is more than 60%. These data are automatically flagged and are not sent to the following processing flow.

The second step consists in calculating an accuracy correction to a reference station using a COCAL box located in its proximity. Since sensors proved to behave consistently among them, following [39] we apply the same accuracy correction to all the other sensors. As abovementioned, given that in the designated area only one value per day is currently available from the reference station, we calculate the difference between that reference value and the daily average value of all the measurements taken by a LCS co-located in the proximity of the reference station. Corrections are inter/extrapolated in all the designated areas by means of the technique described in [39].

An example of the method's results can be seen in Figure 4, where on the left the geographic distribution of LCS measurements before corrections is shown, while on the right corrected values that are more consistent with reference measurements. It is to be noted that this method can be problematic since applying the correction to areas far from where the reference station is located can unpredictably bias the final values. In the case proposed in Figure 4, measurements taken in the village of Opicina (upper part of the map) generally depict rather different conditions from the city center (lower part of the map). Opicina is in fact located uphill and is characterized by a different climatic setting and is less subject to vehicular traffic. No reference station is available in that area so that the only reference measurements available are those taken in the city center. In the example of Figure 4 (left), while the un-corrected data report a polluted city center and a much better situation up-hill, after the corrections (Figure 4 right) the revised air quality degrades notably also in the hills. This could be an artifact that needs careful consideration when interpreting the data.

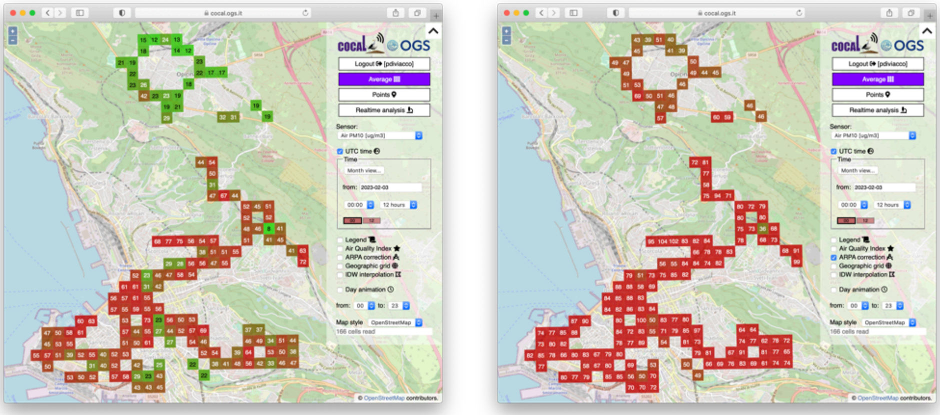


Figure 4. Distribution of raw PM_{10} values (left); distribution of data after correction to improve LCS accuracy (right).

2.2. Deployment on mobile platforms

As mentioned above, sensors were installed on two different platform types, and namely (i) buses and (ii) cars. In both cases, we developed a tailor-made waterproof box that can easily be installed on the platform and where all the acquisition and transmission electronics can be safely protected while air inlets and outlets could effectively bring air samples to the LCS.

Bus deployment has been developed with the key help from the local transportation authority, TPL Trieste Trasporti, that kindly offered to host several COCAL systems. The boxes were installed on the roof of the buses (Figure 5) in a closed compartment with a specific air inlet passing through a syphon in order to prevent rain getting inside the box. Power supply is obtained from the bus using a temporised relay to minimise the impact of the COCAL box on the normal functioning of the vehicles.

Figure 5. Installing a COCAL box on buses

Buses are a very convenient acquisition platform because each unit can be redirected to several routes throughout the day, thereby covering a large portion of the urban area. On the other hand, bus routes tend to follow the main directions of the traffic in a city which may somehow bias the coverage of the designated area.

Cars have several advantages on buses, one being that they generally do not follow predefined routes. This makes cars a good means to provide infills in areas where buses are not available. At the same time, cars introduce other constraints that depend mainly on the volunteer drivers. Issues may arise, in fact, in order to motivate them to cover areas that are not within their daily routines. In our experience, this often meant also using vehicles belonging to our institute.

COCAL boxes for cars have been designed entirely by us and 3D printed autonomously with the help of the ICTP FabLab laboratory (Figure 6). The boxes were conceived to be fully autonomous and less invasive as possible. This forced us to make some design choices; for example, since connecting the boxes to the car's power supply can be problematic, they are powered on batteries only. Autonomy is approximately one full day, although it can be longer depending on the rate of data transmission. Battery recharge can be done in a few hours. Another choice was to avoid taking up space inside the vehicle or in the trunk, so we decided to position the box on the roof. To affix the box to the roof surface we added magnetic plates on the bottom and for further security we decided to install it on cars with roof bars only, to which COCAL boxes are secured using Velcro strips. The air inlet passes through the curved white roof so that it remains dry in case of rain (Figure 3 lower left). The air outlet is located on the back of the box. The roof can easily be removed to access the electronics inside.

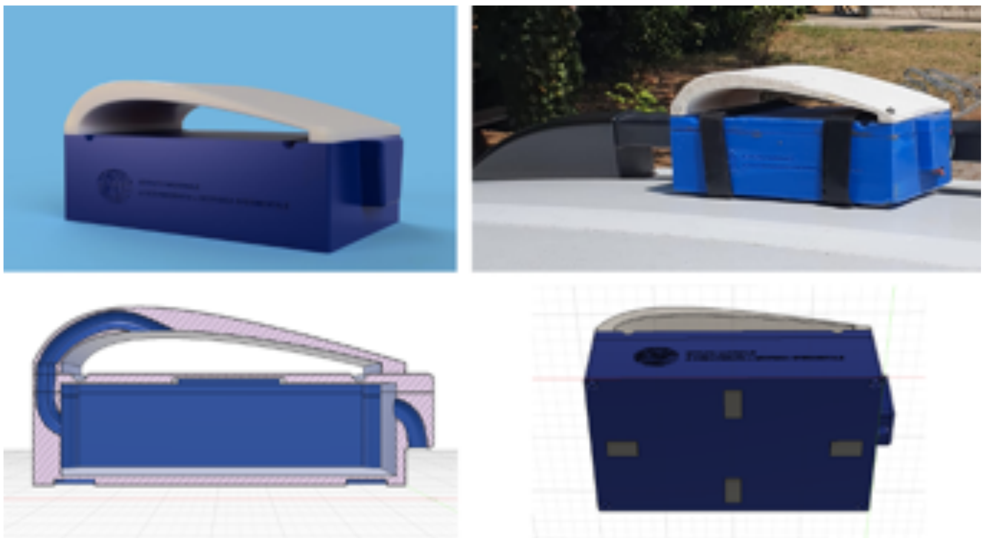


Figure 6. Design and installation of COCAL boxes for cars: Rendering of the box (upper left). Lateral section of the box (lower left): the air inlet passes through the curved white roof to remain dry in case of rain. The air outlet is located on the back of the box. The roof can easily be removed to access the electronics inside. Actual deployment on a car (upper right): boxes are located on car roofs secured with Velcro stripes to roof bars. COCAL boxes have magnetic plates to better adhere to the roofs (lower right).

2.2.1 Limitations of mobile platforms

Besides the already mentioned limitations in accuracy, we were also concerned about the possible effects of the deployment of LCSs on moving platforms. While it is known that platform speed influences the measurement, to our knowledge there is no specific study on this topic since most of the existing literature is based on fixed position deployments. We therefore set up a test, where passing multiple times in the same area at different speeds during a restricted period with stable meteorologic conditions, we collected a large dataset of measurements. Results of the experiment can be seen in Figure 7. These show an inverse relationship between PM and platform speed. Considering how the COCAL box is built, this is probably due to a depression induced by the platform movement on the inlet of the box. This increases with speed, reducing the quantity of air that reaches the detection device and therefore reducing the estimates of PM values. The drift is relatively small and below the precision of sensors for velocities lower than 50 km/h, while higher speed values tend to be more problematic. Since the system has been installed mostly in an area where the speed limit is below 50 Km/h, we can safely say that data collection is not particularly affected by this issue. As a measure of further security, during data processing, measurements associated with a speed higher than 50 km/h are automatically filtered out of the calculations

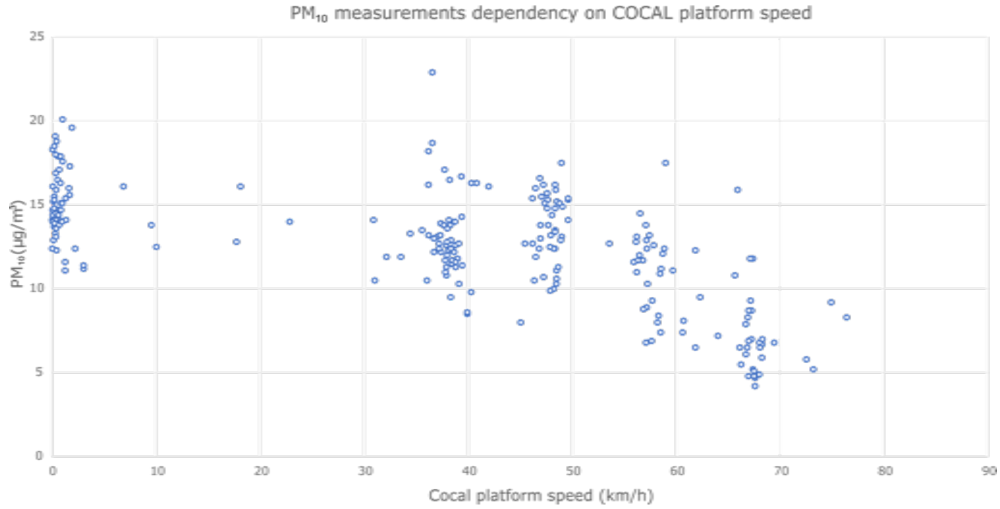


Figure 7. Measurements of PM_{10} concentration as a function of platform speed

2.3. Data acquisition and transmission

The acquisition system (Figure 8) is based on a low-cost ESP32 microcontroller with WiFi and Bluetooth connectivity. We selected a Heltec LoRa 32 v2 board, which has an embedded OLED display and battery charger together with LoRa chip, and Wi-Fi and Bluetooth connectivity. These are used for testing and short distance connectivity while LoRa is used for long distance connectivity [28]. To this, we added GSM and GPS functionalities using an A9G development board, designed by Ai-Thinker, that, with an active SIM card, allows data transmission using the GSM telephonic network where coverage is available. Data transmission using Wi-Fi and GSM stores data directly in an InfluxDB database, while using Lora we rely on The Things Network (TTN) LoRaWAN infrastructure in order to retrieve data transmitted using LoRa and store data into the database. Telegraf, a server-based agent, oversees retrieving data from TTN using MQTT protocol and storing data into the database.

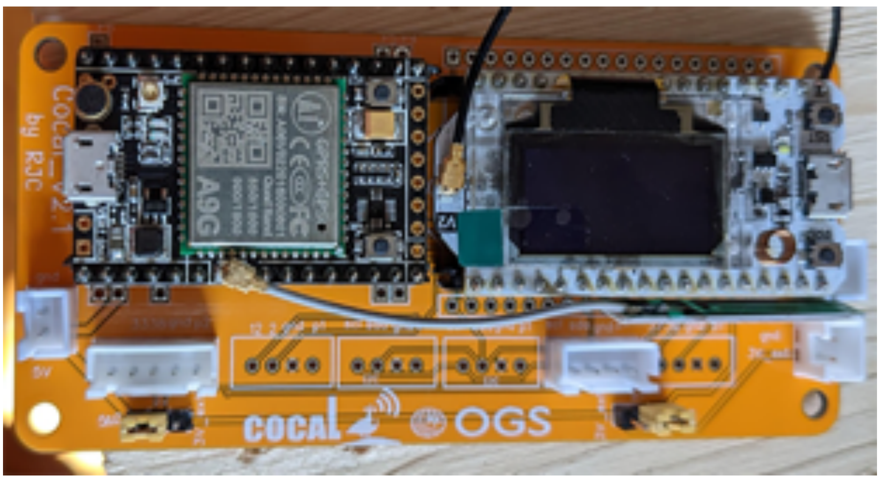


Figure 8. Heltec and A9G board on a PCB designed for COCAL

2.4. Data management

Figure 9 describes the general architecture of the COCAL system. The flux of incoming data transmitted through LoRaWAN flows into an InfluxDB table filled by TTN service. A server script manages to reroute the data into the main InfluxDB time-series tables after proper conversion. The final storage and processing server is based on a Postgres database, with PostGIS extension for dealing with georeferenced objects, geographic projections, and geometric objects like polylines. This storage/processing server (SPS) is built on an open-source architecture: Linux Ubuntu, Apache, PHP, Python and Postgres. It currently manages the database, as well as several scripts responsible for the processing and the web front-end. A PHP script periodically synchronizes the InfluxDB with the Postgres database, inserting in the latter one, every valid measurements from a sensor with a time marker (UTC), WGS84 coordinates, all other GPS info (such as altitude or speed), the type of transmission (e.g. GSM), a device ID, a sensor ID (e.g. atmospheric pressure) and the measured value (e.g. 1007 mBar).

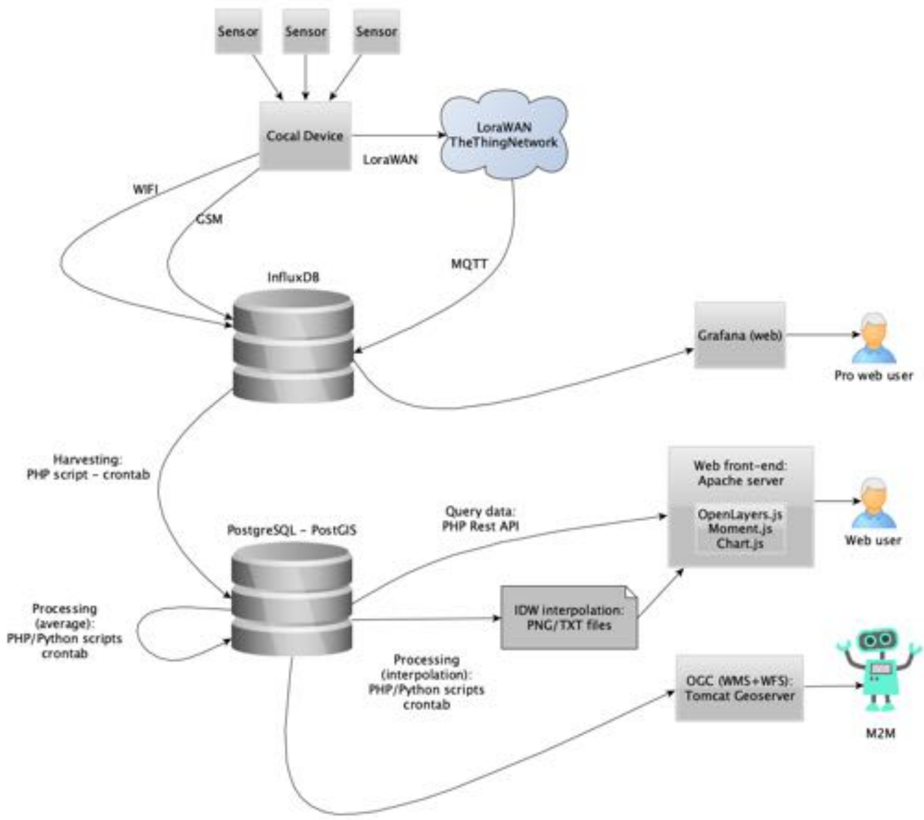


Figure 9. General architecture of the COCAL system

2.5 Data Processing

The SPS performs different activities by means of PHP scripts, which are scheduled with crontab. The most demanding analyses are encoded in Python with its standard libraries such as NumPy, SciPy, Matplotlib or PIL.

All processed products are made available in near real-time and stored permanently for a better performance.

2.5.1 Window averaging

Window averaging is necessary to assimilate the large amount of data acquired by many devices spread across a wide area. As in [35], we define a geographical grid of 200m per 200m wide cells, based on a local projection. In addition, we subdivide the timeline into 1-hour intervals. Every set of data spanning a spatial cell and a time interval is a datacube, including measurements from different devices but mounting the same kind of sensor (e.g. PM_{10}). The choice of a local projection provides good accuracy when the area of interest is limited, and in the case of this work we used WGS84/UTM zone 33N (EPSG:32633). In order to obtain values that are smoother and more representative of the physical phenomenon, reducing the outliers and providing a uniform subdivision of space and time, considering the good results obtained in [26], we adopted a similar approach by averaging data (calculating the median) over space and time datacubes. Larger time intervals (for example 2, 3, 4 or 8 hours) can be analyzed selecting the specific datacube. These are processed once a day and made available the next day. A discussion on the advantages of window averaging and the shape of the cells can also be found in [39].

Every averaged data cube is stored into the database, marked with start time, end time and a polyline describing the square cell.

2.5.2. Correction of LCS data

As mentioned in section 2.1.2.2, the accuracy of LCS can be problematic. The technologies used within these sensors, the rate of sampling together with the effects of environmental parameters such as the RH, often induce drifts in the LCS measurements. In section 2.1.2.3, we introduced a pragmatic method that can mitigate such effects. This is based on applying a correction value to the LCS measurements that is calculated daily as the difference between the value provided by a reference station and a fixed LCS located in the vicinity of the reference station. The correction is applied server side one day after the LCS data are actually collected since the reference value is available only with such delay. The resulting grid of data is then made available to the web portal.

2.5.3. Interpolation and contouring.

In order to provide a more intuitive insight into the measured phenomenon, interpolation is a useful tool. Following [39], there are many aspects to take into account when spatial interpolation is applied:

(i) The accuracy of the method and how far the interpolated values from the samples are still meaningful, i.e. a consideration on "extrapolation". This issue can be partially solved by the definition of an area of interpolation, like the bounding box or (better) the convex hull of the samples as first approximation.

(ii) The computational complexity and the relative speed of the interpolation method, which must comply with the near real-time requirement. In our implementation, we chose a very quick and sufficiently accurate method, the Inverse Distance Weighting (IDW). IDW interpolation is defined as follows. Assuming that $\{x_1 \ldots x_N\}$ are the interpolating points (samples) and $\boldsymbol{x}$ a generic point, the interpolant is:

$$f_{IDW}(\boldsymbol{x}) = \frac{\sum_{k=1}^{N} w_k(\boldsymbol{x})\, f(\boldsymbol{x}_k)}{\sum_{k=1}^{N} w_k(\boldsymbol{x})} \quad (1)$$

Where the weights are:

$$w_k(x) = \| x - x_k \|^{-p} \quad (2)$$

(iii) In addition, there is an epistemological aspect to be further considered: all the processing is automatic and cannot be assisted by a human intervention. This fact excludes algorithms like Kriging, which involve many discretional models and parameters.

An excellent alternative solution is Natural Neighbours Interpolation (Sibson's method), which is based only on the geometrical properties of the dataset, and is approximately ten times slower than IDW [50]. Lastly, it is necessary to define the levels for the contouring in some adaptive way, to improve the readability and also the colour map for the interpolation, which must be coherent with all other data visualisations. The interpolation/contouring is implemented on the SPS with a Python script that reads the averaged values, applies the IDW method (with exponent p=3), generates the contour and produces a transparent PNG image with a small text file for the georeferentiation. The image is clipped around the convex hull of the dataset, excluding the external area.

2.5.4. Near-real time web-based visualization.

The visualisation of environmental data is a topic that raises several questions: Are our data time-series or spatial distributions? How to represent time-varying phenomena? Which colours and graphic patterns are more effective and representative? To what extent does the computation have an impact on near-real time web interaction? There are many different answers of course and much research involving mathematical, computational or psychological aspects (see for example [39]).

We implemented a set of visualisations in the web front-end, which allows the end user to browse through spatial and temporal coordinates and to select and analyse both single acquisitions and averaged maps (Figure 10).

All services are available at the web portal https://cocal.ogs.it.

The web interface allows an easy navigation through the datasets, by means of a simple window (Figure 10 left) where the user can select the single device, the acquisition sensor, the time interval and many different options. The time selection can be done in local time or in UTC and a simplified view of the day shows the density of available data as shades of red, providing a one-click access to time selection.

A calendar (month view button) shows the data density day by day, by using the same principle. The acquisitions of a single device are represented as an interactive chart of the time-series (a) or as a collection of connected points on a map (b). In the latter case, an arrow shape can show the GPS direction and the point colours are mapped to the measure scale and corresponding legend.

The averaged data (on a rectangular grid) are represented as colored and labelled polygons (c). The interpolated data use the same colour coding but are represented as continuous within the data convex hull, with superimposed contour lines (d). All graphic elements are responsive, showing all data details.

Additionally, we implemented a functionality that allows the user to follow cumulative data as an animation, cycling from a starting to an ending hour, in order to dynamically represent the temporal evolution of each parameter.

2.6.5. Advanced analysis

The adoption of UTM33N as the map projection is a disadvantage when the acquisitions are beyond the limits of the range from 12ºE to 18ºE. Moreover, the implemented processing mechanism, which computes the averages periodically (in background), is very efficient for a quick response and a fluid user experience, but on the down side is rather fixed and rigid. A more flexible interface for data analysis has been tested, based on the global map projection "Web Mercator" (EPSG:3857, Pseudo-Mercator / Spherical Mercator). This spherical projection is used by most GIS systems like Google Maps, Bing, ESRI, etc., has an increasing distortion at high latitudes and is not conformal, but is the de facto standard for web applications and allows global coverage for the processing. The web page shown in Figure 11 provides a wider range of query parameters and builds "on the fly" an analysis (windowed average or IDW interpolation). The computation is restricted to the visible bounding box and requires some computational time, favoring the extended query flexibility.

2.6.6. Open and FAIR data access.

According to the FAIR principles, data must be Findable, Accessible, Interoperable and Reusable. COCAL deploys different protocols and implementations aiming to provide Open and FAIR data in accordance with well-established and official standards. In order to achieve discoverability, the initiative handles standard ISO 19115-3 metadata profile through the Geonetwork catalogue application. To ensure interoperability, such as machine to machine data flows, data harvesting or archive federations, COCAL geospatial database is compliant to OGC (Open Geospatial Consortium) standards, deployed as Web HTTP services: (i) WMS (Web Map Service), which provides georeferenced map images of the requested features; (ii) WFS (Web Feature Service), which provides detailed and fine-grained information about features or general capabilities of the dataset in a structured text format (XML, JSON, etc.). All OGC services are implemented on a Geoserver platform, an open-source Apache Tomcat extension linked to the main database. In order to achieve accessibility, data products are fully open and accessible, while download of raw data in CSV format is available on the COCAL web portal, after authentication with a trusted account.

Figure 10. COCAL web-based GUI (left) and near real-time visualizations: a) time series; b) single acquisitions; c) near real-time window averaging; d) interpolation and contour.

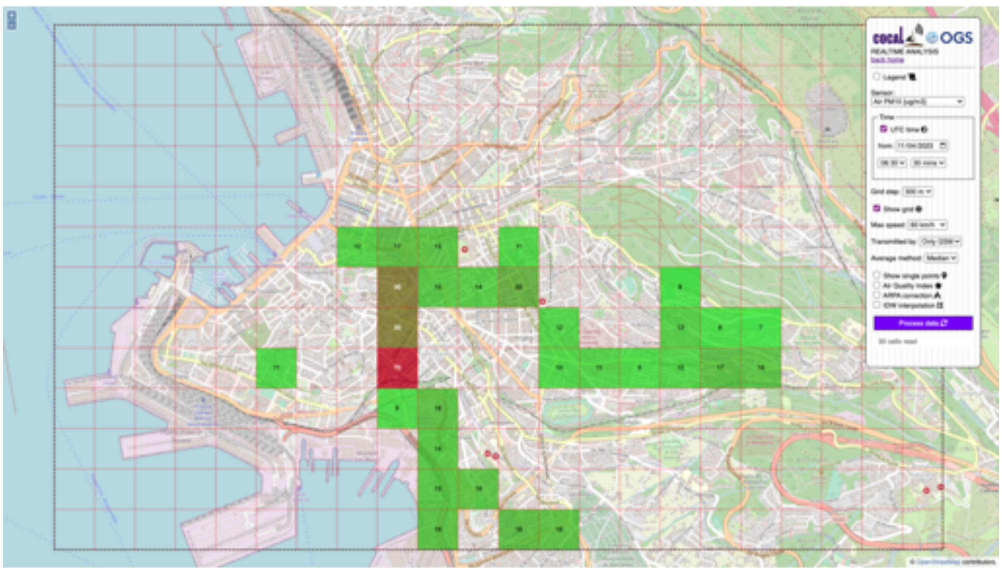


Figure 11. COCAL web “real time analysis” with advanced queries.

3. Results

The technology behind COCAL has been under constant development since 2020. During an initial phase, trials took place using multiple simultaneous acquisition and transmission platforms mounted on vehicles operated by our institution. This allowed us to test extensively the system, its scalability, precision and accuracy during specific targeted surveys. Once the system was finalized, we were able to deploy a fully operative system on vehicles of the local transportation authority (Trieste Trasporti) and on some voluntary cars. COCAL entered into service in February 2021 and has been fully operative ever since.

Up to April 2023, the system acquired and processed a remarkable amount of data, both "points" (single measurements) and "cells" (averaged results). In Table 2, it is possible to see the approximated number of records per year during the period from January 2020 to April, 2023.

Table 2. Approximated number of records per year.

Record type	Total count	Year	Count
Points	53M	2020	3.8M
		2021	13M
		2022	27M
		2023 (partial)	9.2M
Cells	18M	2020	0.9M
		2021	5.8M
		2022	9.4M
		2023 (partial)	1.8M

Data is fully public and can be accessed using standard procedures from the COCAL web portal (https://cocal.ogs.it). The main results of the work are on one hand the COCAL system itself, which has proved to be efficient, robust, easy to install and maintain, allowing a very high throughput of environmental data that strongly support the paradigm of low-cost participative systems in monitoring the environment. On the other hand, a very important result is the availability of a very large quantity of environmental measurements, allowing to significantly increase the spatial and time coverage of the distribution of air pollutants in the designated area. Initial analysis of the dataset acquired enabled identifying several interesting features of the air quality in the area of Trieste. Some of these observations have already been published in the papers mentioned above.

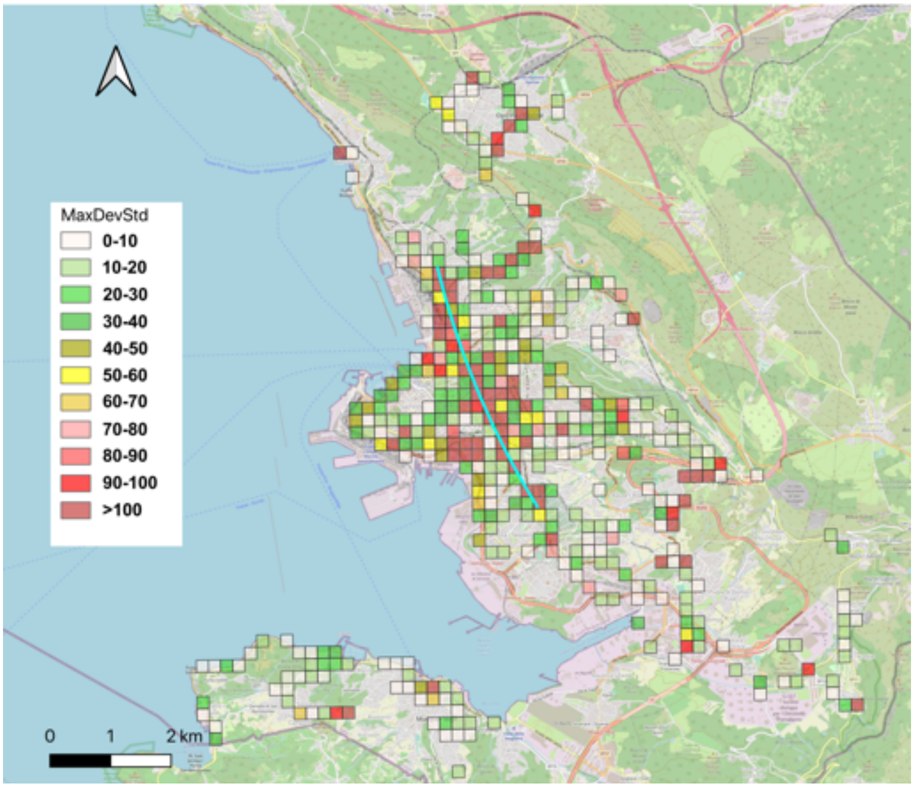


Figure 12. Map of maximum standard deviation of COCAL PM_{10} measurements in the designated area. In light blue, the alignment of the higher values correlates with the main traffic direction of the city.

Figure 12 shows the geographic distribution of the standard deviation of all measurements done by the COCAL systems in the designated area. For every possible grid cell we consider all the PM measurements during the whole considered period (...2021...2022). Acquisitions within each cell are divided into time windows of 1 hour and some statistical parameters are calculated for every window, such as the number of samples, average, median and standard deviation. Eventually, the maximum standard deviation over the period is calculated for every cell.

This map should be interpreted with some caution. On one hand, the standard deviation could provide an idea of the amount of variation of measurements during a specific period of time. Where the standard deviation is high, this could mean that higher levels of PM have been recorded in that area compared to areas where the standard deviation is lower. On the other hand, there is a risk that if a biased coverage of data has been used then the standard deviation could also be biased. Indeed, in Figure 12, it is easy to identify a NW-SE alignment (highlighted by the light blue line) where high values of the standard deviations seem to gather. This correlated almost perfectly with the main traffic direction of the city. It could be tempting to associate this trend with the traffic, concluding that those areas are the most polluted in the city. We think that some caution should be taken, because this direction also corresponds to the more frequently followed routes of the buses used in the COCAL initiative. Most of the data, in fact, have been acquired in those areas so that in comparison to other areas measurements done in the former could have had the opportunity to detect all pollution events, while measurements done in other areas may just have overlooked them.

An interesting and surprising result we obtained is related to the possible deterioration of the sensors after long usage. After approximately 2 years of activity, we decided to substitute the hardware of the deployed systems and discovered that they practically had not accumulated any dirt inside and that their precision remained (considering the intrinsic limitations of the LCS) almost unaltered.

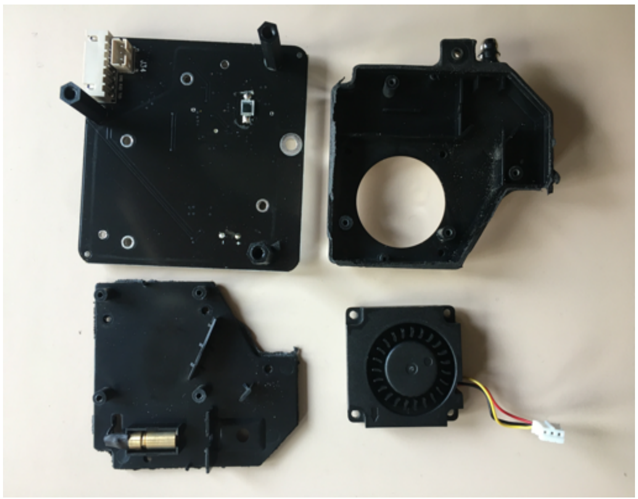


Figure 13. An LCS case opened after almost two years of activity. The interiors show only very small quantities of dirt.

5. Conclusions and future work

This work describes a crowdsensing based air monitoring system following all the technological segments of a path that starts from the actual measurement using LCS, to data transmission, to processing and FAIR compliant web-based representation and access to the reconstructed data products. The main conclusion of this work is therefore that the implementation of all segments of the system can be achieved using low-cost and open-source technology only. At the same time, acquisition of data does not need trained personnel but can be done with the help of volunteers and especially of the local transportation authorities. The results of the experience we propose here, suggest that such systems can be trustworthy from the point of view of the precision of measurements, while it is necessary to rely on a reference value to correct the deviation of measurements due to the intrinsic limitations of the LCSs. Currently, within the proposed system, this correction is calculated daily, since the reference values are made available by the local Environmental Agency only as daily averages and one day after the actual measurements took place. We demonstrated that, in some cases, this can be problematic and that, when available, corrections should be calculated with a higher temporal resolution. Other limitations of the system were taken into consideration such as the speed of the acquisition vehicle. We showed that this speed should not exceed 60 km/h otherwise, the variation in air pressure could bias the measurements. We also demonstrated that after almost two years of continuous field operations, the amount of dirt accumulated within the acquisition box in the case of the designated area was minimal. This, of course, can depend on the levels of pollution of the specific cases where the system is applied.

The amount of data acquired raised important questions from the viewpoint of the ITC system to be used. We understood that, for example, a modular approach in separating each activity on different virtual machines helps considerably in monitoring the performances of the system and in understanding where it may be necessary to increase the dedicated resources.

After two years of testing the system with five COCAL boxes installed on local transportation authority buses, we understood that since this means of transportation can often be under maintenance or rerouted, five units of vectors is the minimum set of installation for a city of approximately 200.000 inhabitants and an area of approximately 100.000 square meters. In this perspective, much depends on the actual urban configuration. In fact, while installations on cars can cover the urban area almost randomly, buses follow the bus line distribution which is generally concentrated in specific areas while neglecting others. Resulting data can be biases and limit the reconstruction of the distribution of air quality.

In the next months, the number of COCAL installations on the public bus network of Trieste will be doubled in order to achieve a broader and more homogeneous coverage of the city.

The designated area being a coastal area, the system developed so far allows reconstructing the on-land area only. This does not allow studying the phenomena in depth related to the movements of polluted air masses due to sea breezes. In this perspective, an extension of the system is planned at sea using volunteers' recreational sailing ships, while the system will also be installed on a certain number of sea buoys managed by OGS.

Author Contributions: Conceptualization, P.D; methodology, P.D., M.I and R.C..; software, M.I., R.C..; validation, P.D., M.I., R.C, A.V and N.P.; resources, R.C., N.P., A.V; data management, M.I, N.P.; writing—original draft preparation, P.D, M.I, R.C.; writing—review and editing, P.D.; visualization, M.I..; supervision, P.D.; project administration, P.D. All authors have read and agreed to the published version of the manuscript

Funding: This research received no external funding.

Data Availability Statement: All the data used in this work are freely and openly available at the web site https://COCAL.ogs.it accessed on 3 May 2023.

Acknowledgments: The authors would like to thank Trieste Trasporti TPL, the local transportation authority of Trieste, for its support, without which this work would not have been possible. And, in particular, we would also like to thank, for their great help, Giuseppe Zottis, Fabrizio Godinich, and Fabio Vidotto. We are also very grateful to the ICTP FabLAb for their effective help in designing and 3D printing the COCAL boxes.

Conflicts of Interest: "The authors declare no conflict of interest".

References

[1] 'Air pollution'. https://www.who.int/health-topics/air-pollution (accessed Feb. 20, 2023).

[2] A. A. Almetwally, M. Bin-Jumah, and A. A. Allam, 'Ambient air pollution and its influence on human health and welfare: an overview', *Environ. Sci. Pollut. Res.*, vol. 27, no. 20, pp. 24815–24830, Jul. 2020, doi: 10.1007/s11356-020-09042-2.

[3] F. Di Turo *et al.*, 'Impacts of air pollution on cultural heritage corrosion at European level: What has been achieved and what are the future scenarios', *Environ. Pollut.*, vol. 218, pp. 586–594, Nov. 2016, doi: 10.1016/j.envpol.2016.07.042.

[4] 'Cachier: Aerosol characterization and sources in... - Google Scholar'. https://scholar.google.com/scholar_lookup?title=Aerosol%20characterization%20and%20sources%20in%20different%20European%20urban%20atmospheres%3A%20Paris%2C%20Seville%2C%20Florence%20and%20Milan&publication_year=2004&author=H.%20Cachier&author=R.%20Sarda-Est%C3%A8ve&author=K.%20Oikonomou&author=J.%20Sciare&author=A.%20Bonazza&author=C.%20Sabbioni&author=M.%20Greco&author=J.%20Reyes&author=B.%20Hermosin&author=C.%20Saiz-Jimenez (accessed Feb. 20, 2023).

[5] O. US EPA, 'Clean Air Act Text', May 29, 2015. https://www.epa.gov/clean-air-act-overview/clean-air-act-text (accessed Feb. 21, 2023).

[6] *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*, vol. 152. 2008. Accessed: Feb. 21, 2023. [Online]. Available: http://data.europa.eu/eli/dir/2008/50/oj/eng

[7] 'European city air quality viewer — European Environment Agency'. https://www.eea.europa.eu/themes/air/urban-air-quality/european-city-air-quality-viewer (accessed Feb. 22, 2023).

[8] O. US EPA, 'Interactive Map of Air Quality Monitors', Aug. 17, 2016. https://www.epa.gov/outdoor-air-quality-data/interactive-map-air-quality-monitors (accessed Feb. 22, 2023).

[9] M. Strzelecki and K. Stezycki, '"Burn everything": Poland chokes on the smog of war', *Reuters*, Dec. 08, 2022. Accessed: Feb. 22, 2023. [Online]. Available: https://www.reuters.com/world/europe/burn-everything-poland-chokes-smog-war-2022-12-08/

[10] EPA, 'AP-42: Compilation of Air Emissions Factors'. EPA. Accessed: Aug. 12, 2022. [Online]. Available: https://www3.epa.gov/ttnchie1/ap42/ch01/

[11] D. Pudasainee, V. Kurian, and R. Gupta, '2 - Coal: Past, Present, and Future Sustainable Use', in *Future Energy (Third Edition)*, T. M. Letcher, Ed., Elsevier, 2020, pp. 21–48. doi: 10.1016/B978-0-08-102886-5.00002-5.

[12] D. Singh, M. Dahiya, R. Kumar, and C. Nanda, 'Sensors and systems for air quality assessment monitoring and management: A review', *J. Environ. Manage.*, vol. 289, p. 112510, Jul. 2021, doi: 10.1016/j.jenvman.2021.112510.

[13] G. Bonafè, F. Montanari, and F. Stel, 'Air quality in Trieste, Italy - A hybrid Eulerian-Lagrangian-statistical approach to evaluate air quality in a mixed residential-industrial environment', p. 37.

[14] Irwin, A., *Citizen Science: A Study of People, Expertise and Sustainable Development. Lond*. London: Routledge.

[15] F. Froeling *et al.*, 'Narrative review of citizen science in environmental epidemiology: Setting the stage for co-created research projects in environmental epidemiology', *Environ. Int.*, vol. 152, p. 106470, Jul. 2021, doi: 10.1016/j.envint.2021.106470.

[16] D. Fraisl *et al.*, 'Mapping citizen science contributions to the UN sustainable development goals', *Sustain. Sci.*, vol. 15, no. 6, pp. 1735–1751, Nov. 2020, doi: 10.1007/s11625-020-00833-7.

[17] C. Stewart, G. Labrèche, and D. L. González, 'A Pilot Study on Remote Sensing and Citizen Science for Archaeological Prospection', *Remote Sens.*, vol. 12, no. 17, Art. no. 17, Jan. 2020, doi: 10.3390/rs12172795.

[18] D. Kasperowski and T. Hillman, 'The epistemic culture in an online citizen science project: Programs, antiprograms and epistemic subjects', *Soc. Stud. Sci.*, vol. 48, no. 4, pp. 564–588, Aug. 2018, doi: 10.1177/0306312718778806.

[19] C. Kullenberg and D. Kasperowski, 'What Is Citizen Science? – A Scientometric Meta-Analysis', *PLOS ONE*, vol. 11, no. 1, p. e0147152, Jan. 2016, doi: 10.1371/journal.pone.0147152.

[20] F. Ward *et al.*, 'Engaging communities in addressing air quality: a scoping review', *Environ. Health*, vol. 21, no. 1, p. 89, Sep. 2022, doi: 10.1186/s12940-022-00896-2.

[21] A. Höhne *et al.*, 'Assessing the Spatial Distribution of NO2 and Influencing Factors in Urban Areas—Passive Sampling in a Citizen Science Project in Berlin, Germany', *Atmosphere*, vol. 14, no. 2, Art. no. 2, Feb. 2023, doi: 10.3390/atmos14020360.

[22] S. De Craemer, J. Vercauteren, F. Fierens, W. Lefebvre, and F. J. R. Meysman, 'Using Large-Scale NO2 Data from Citizen Science for Air-Quality Compliance and Policy Support', *Environ. Sci. Technol.*, vol. 54, no. 18, pp. 11070–11078, Sep. 2020, doi: 10.1021/acs.est.0c02436.

[23] J. A. Ellenburg *et al.*, 'Global Ozone (GO3) Project and AQTreks: Use of evolving technologies by students and citizen scientists to monitor air pollutants', *Atmospheric Environ. X*, vol. 4, p. 100048, Oct. 2019, doi: 10.1016/j.aeaoa.2019.100048.

[24] A. Baldelli, 'Evaluation of a low-cost multi-channel monitor for indoor air quality through a novel, low-cost, and reproducible platform', *Meas. Sens.*, vol. 17, p. 100059, Oct. 2021, doi: 10.1016/j.measen.2021.100059.

[25] D. Bousiotis, L.-N. S. Alconcel, D. C. S. Beddows, R. M. Harrison, and F. D. Pope, 'Monitoring and apportioning sources of indoor air quality using low-cost particulate matter sensors', *Environ. Int.*, vol. 174, p. 107907, Apr. 2023, doi: 10.1016/j.envint.2023.107907.

[26] G. Yang, Y. Zhou, and B. Yan, 'Contribution of influential factors on PM2.5 concentrations in classrooms of a primary school in North China: A machine discovery approach', *Energy Build.*, vol. 283, p. 112787, Mar. 2023, doi: 10.1016/j.enbuild.2023.112787.

[27] M. Zusman *et al.*, 'Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study', *Environ. Int.*, vol. 134, p. 105329, Jan. 2020, doi: 10.1016/j.envint.2019.105329.

[28] H. J. Johnston *et al.*, 'How Harmful Is Particulate Matter Emitted from Biomass Burning? A Thailand Perspective', *Curr. Pollut. Rep.*, vol. 5, no. 4, pp. 353–377, Dec. 2019, doi: 10.1007/s40726-019-00125-4.

[29] R. K. Ganti, F. Ye, and H. Lei, 'Mobile crowdsensing: current state and future challenges', *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, Nov. 2011, doi: 10.1109/MCOM.2011.6069707.

[30] D. Fraisl *et al.*, 'Citizen science in environmental and ecological sciences', *Nat. Rev. Methods Primer*, vol. 2, no. 1, Art. no. 1, Aug. 2022, doi: 10.1038/s43586-022-00144-4.

[31] U. Wehn *et al.*, 'Impact assessment of citizen science: state of the art and guiding principles for a consolidated approach', *Sustain. Sci.*, vol. 16, no. 5, pp. 1683–1699, Sep. 2021, doi: 10.1007/s11625-021-00959-2.

[32] M. Aristeidou and C. Herodotou, 'Online citizen science: A systematic review of effects on learning and scientific literacy', *Citiz. Sci. Theory Pract.*, vol. 5, no. 1, 2020, doi: 10.5334/cstp.224.

[33] T. M. Luger, A. B. Hamilton, and G. True, 'Measuring Community-Engaged Research Contexts, Processes, and Outcomes: A Mapping Review', *Milbank Q.*, vol. 98, no. 2, pp. 493–553, 2020, doi: 10.1111/1468-0009.12458.

[34] A. J. Friedman, 'Report from a National Science Foundation Workshop'.

[35] P. Diviacco *et al.*, 'MaDCrow, a Citizen Science Infrastructure to Monitor Water Quality in the Gulf of Trieste (North Adriatic Sea)', *Front. Mar. Sci.*, vol. 8, p. 619898, Jul. 2021, doi: 10.3389/fmars.2021.619898.

[36] P. Diviacco *et al.*, 'Citizen science and crowdsourcing in the field of marine scientific research — the MaDCrow project', *J. Sci. Commun.*, vol. 20, no. 06, p. A09, Oct. 2021, doi: 10.22323/2.20060209.

[37] P. Diviacco, M. Iurcev, R. J. Carbajales, and N. Potleca, 'First Results of the Application of a Citizen Science-Based Mobile Monitoring System to the Study of Household Heating Emissions', *Atmosphere*, vol. 13, no. 10, Art. no. 10, Oct. 2022, doi: 10.3390/atmos13101689.

[38] P. Diviacco *et al.*, 'Monitoring Air Quality in Urban Areas Using a Vehicle Sensor Network (VSN) Crowdsensing Paradigm', *Remote Sens.*, vol. 14, no. 21, Art. no. 21, Jan. 2022, doi: 10.3390/rs14215576.

[39] Iurcev, M., Pettenati, F., and Diviacco, P., 'Improved automated methods for near real-time mapping - application in the environmental domain', *Bull. Geophys. Oceanogr.*, vol. 62, no. 3, pp. 427–454.

[40] K. Horvath, S. Ivatek-Šahdan, B. Ivančan-Picek, and V. Grubišić, 'Evolution and Structure of Two Severe Cyclonic Bora Events: Contrast between the Northern and Southern Adriatic', *Weather Forecast.*, vol. 24, no. 4, pp. 946–964, Aug. 2009, doi: 10.1175/2009WAF2222174.1.

[41] M. Orlic, B. Penzar, and I. Penzar, 'Adriatic Sea and Land Breezes: Clockwise Versus Anticlockwise Rotation', *J. Appl. Meteorol.*, vol. 27, pp. 675–679, Apr. 1988, doi: 10.1175/1520-0450(1988)027<0675:ASALBC>2.0.CO;2.

[42] F. M. J. Bulot *et al.*, 'Laboratory Comparison of Low-Cost Particulate Matter Sensors to Measure Transient Events of Pollution', *Sensors*, vol. 20, no. 8, Art. no. 8, Jan. 2020, doi: 10.3390/s20082219.

[43] N. Zimmerman, 'Tutorial: Guidelines for implementing low-cost sensor networks for aerosol monitoring', *J. Aerosol Sci.*, vol. 159, p. 105872, Jan. 2022, doi: 10.1016/j.jaerosci.2021.105872.

[44] L. Liang, 'Calibrating low-cost sensors for ambient air monitoring: Techniques, trends, and challenges', *Environ. Res.*, vol. 197, p. 111163, Jun. 2021, doi: 10.1016/j.envres.2021.111163.

[45] W. Y. Yi, K. M. Lo, T. Mak, K. S. Leung, Y. Leung, and M. L. Meng, 'A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems', *Sensors*, vol. 15, no. 12, Art. no. 12, Dec. 2015, doi: 10.3390/s151229859.

[46] M. I. Mead *et al.*, 'The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks', *Atmos. Environ.*, vol. 70, pp. 186–203, May 2013, doi: 10.1016/j.atmosenv.2012.11.060.

[47] L. R. Crilley *et al.*, 'Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring', *Atmospheric Meas. Tech.*, vol. 11, no. 2, pp. 709–720, Feb. 2018, doi: 10.5194/amt-11-709-2018.

[48] L. R. Crilley *et al.*, 'Effect of aerosol composition on the performance of low-cost optical particle counter correction factors', *Atmospheric Meas. Tech.*, vol. 13, no. 3, pp. 1181–1193, Mar. 2020, doi: 10.5194/amt-13-1181-2020.

[49] W. C. Malm, D. E. Day, S. M. Kreidenweis, J. L. Collett, and T. Lee, 'Humidity-dependent optical properties of fine particles during the Big Bend Regional Aerosol and Visibility Observational Study', *J. Geophys. Res. Atmospheres*, vol. 108, no. D9, 2003, doi: 10.1029/2002JD002998.

[50] M. Sathiyanarayanan, V. Varadarajan, and K. V. Pradeep, 'Visual analytics on spatial time series for environmental data', *Int. J. Recent Technol. Eng.*, vol. 8, no. 1C2, pp. 1173–1181, 2019.

Paolo Diviacco
GEOPHYSICS
National Institute of Oceanography and Applied Geophysics
Borgo Grotta Gigante 42/c
Sgonico (Trieste) 34010
Italy
VAT: IT00055590327

INVOICE

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel.: +41 61 683 77 34
E-Mail: billing@mdpi.com
Website: www.mdpi.com
VAT nr. CHE-115.694.943

Date of Invoice:	20 June 2023
Manuscript ID:	processes-2411337
Invoice Number:	2411337
Your Order:	by e-mail (pdiviacco@ogs.it) on 6 May 2023
Article Title:	"Design and Implementation of a crowdsensing based air quality monitoring open and FAIR data infrastructure"
Name of co-authors:	Paolo Diviacco, Massimiliano Iurcev, Rodrigo Jose Carbajales, Alberto Viola and Nikolas Potleca Additional Author Information
Terms of payment:	10 days
Due Date:	30 June 2023
VAT:	VAT reversed
License:	CC BY

Description	Currency	Amount
Article Processing Charges	EUR	2 040.13
Subtotal without VAT	EUR	2 040.13
VAT (0%)	EUR	0.00
Total with VAT	**EUR**	**2 040.13**

Accepted Payment Methods

1. Online Payment by Credit Card in Euros (EUR)

Please visit https://payment.mdpi.com/2348781 to pay by credit card. We accept payments in Euros (EUR) made through VISA, MasterCard, Maestro, American Express, Diners Club and Discover.

2. Paypal in Euros (EUR)

Please visit https://payment.mdpi.com/payment/paypal and enter the payment details. Note that the fee for using Paypal is 5% of the invoiced amount.

3. Wire Transfer in Euros (EUR)

Important: **Please provide the Manuscript ID (processes-2411337) when transferring the payment**

Payment in EUR must be made by wire transfer to the MDPI bank account. Banks fees must be paid by the customer for both payer and payee so that MDPI can receive the full invoiced amount.

IBAN: CH14 0483 5160 4356 5200 0
Beneficiary's Name: MDPI AG
Beneficiary's Address: St. Alban-Anlage 66, CH-4052 Basel, Switzerland
Bank Account Number (EUR, Euros Account for MDPI): 0060-1604356-52
Bank Name: Credit Suisse
Bank Address: Credit Suisse, St. Alban-Graben 1-3, Postfach 2560, CH-4002 Basel, Schweiz
SWIFT code (Wire Transfer Address): CRESCHZZ80A
Clearing number: 4835

For detailed payment instruction, or for more alternative payment methods, visit the website at https://www.mdpi.com/about/payment.

Invoiced Amount in CHF: 2 000.00
Exchange rate applied to this invoice 20 June 2023: 0.98033 EUR/CHF

Thank you for choosing MDPI.