Mobile Sensor Networks for Air Quality Monitoring in Urban Settings

PAR

Adrian ARFIRE

EPFL

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2016

Acknowledgements

T HE accomplishment of this work would not have been possible without the help and support of many people. First of all, I would like to thank my advisor, Alcherio Martinoli, who gave me the opportunity to pursue this research and provided me with constant support throughout my studies. To his great credit, he has managed to build and maintain a lab culture that encourages members to feel part of a team, sharing their specific expertise while being open to learn from each other, even when working on quite different research topics. It has been a privilege to work on my thesis in this environment and I thank him for allowing me to be a part of this lab.

I would also like to thank my thesis committee – Dr. Alireza Karimi, Dr. Guillermo Barrenetxea, Prof. Achim Lilienthal and Prof. Satoshi Takahama - for taking the time to review this manuscript and for providing valuable feedback.

I am thankful to all the people that contributed to the research presented in this thesis. I would particularly like to thank my close collaborators and co-authors: Ali Marjovi, an expert in mobile chemical sensing and mapping, whose advice and support had a significant impact on the success of my research; Emmanuel Droz, an experienced electronics engineer with highly developed mechanical design and fabrication skills, without whose support our mobile sensor network could not have been built; Alexander Bahr, a very skilled scientist and engineer whose experience with real-world deployments proved invaluable during the key stages of system development and deployment; Julien Eberle, a highly talented computer scientist and an exceptional team player who did an excellent job in adapting the Global Sensor Networks middleware to the specific constrains of our project. Other collaborators who contributed to the success of this work, through their student projects or civil service work, and whom I would like to thank are: Christophe Paccolat, for his mechanical design work; Loïc Frund, for his implementation of a street assignment algorithm; Nicolas Bigler, for his work on data quality and calibration; Kévin André, Alex Bonvin, Thomas Coral, Jonathan Giezendanner, Fabrizzio Gonzales and Fanny Jeanneret-Grosjean, who all contributed to creating a baseline in high resolution air pollution mapping.

I would like to thank all my current and former lab colleagues. In particular I am very thankful to: Maria Boberg, with whom I shared the very challenging experience of conducting wind tunnel experiments; Miloš Vasić, with whom I had the pleasure of collaborating on an electric

Acknowledgements

car platform; Ezequiel Di Mario, Chris Evans and José Nuno Pereira, who where there for most of the journey; the senior colleagues from whose experience I learned a lot: Thomas Lochmatter, Sven Gowal, and, especially, Amanda Prorok, to whom I am deeply grateful for having introduced me to the lab. I am also thankful to Corinne Farquharson and Denis Rochat, who, in the background, made sure that the lab was always working smoothly.

I am grateful to all the OpenSense project team members and external partners. I would like to thank the wonderful people of the Lausanne public transportation company, in particular: Thierry Bignardi, Pierre Gaillard and Willy Grandjean. I am also thankful to Céline Beluche, from SGX Sensortech, for her support in chemical sensor characterization. I am grateful to Naneos founder Martin Fierz for sharing his experience and expertise in particulate matter instrumentation.

I would like to thank my family: my brother, Cristian, a highly skilled micro-engineer, on whom I could always count for both technical advice and moral support, and my parents, who have brought me up with a deep respect for education and the pursuit of knowledge.

Finally, my deepest gratitude goes to my wife, Andreea, whose love and support during these years have made the tougher parts of this experience easier to bare, and the satisfying parts all the more enjoyable.

Lausanne, 4 May 2016   Adrian Arfire

ii

Abstract

W ITH urban populations rapidly increasing and millions of deaths being attributed annually to air pollution, there is a critical need for a deeper understanding of urban air quality. The locality of urban emissions sources, and the specific topography of cities lead to a very heterogeneous pollutant concentration landscape, the details of which cannot be captured by traditional monitoring stations. Although highly accurate, these systems are large, heavy and very expensive, which leads to a very sparse distribution of measurements (e.g., in Switzerland only 16 such stations are available for the whole country).

Mobile wireless sensor networks hold the potential to allow a paradigm shift in our understanding of urban air pollution, through a significant augmentation of the spatial resolution of measurements. The road to achieving reliable high quality information from these type of systems is, however, full of challenges. These start already from the system design level, as the task of developing robust large-scale mobile sensing networks for continuous urban monitoring is arduous in itself. The limitations of existing sensor technology is another important source of hard problems. Low-cost chemical sensors, the typical candidates for this type of applications, suffer from a large list of issues that make their use in a mobile scenario non-trivial. These include: instability (temporal drift), cross-sensitivity (to other chemicals or environmental parameters), low signal-to-noise ratios, and slow dynamic response. The latter problem, in particular, is a tough challenge when considering a mobile scenario, since it leads to significant measurement distortion. The question of achieving and maintaining the calibration of chemical sensors, throughout the deployment lifetime, is another essential issue that derives from their typical instability. Finally, the development of appropriate modeling techniques that would enable us to generate high-resolution pollution maps based on mobile sensor network data is a highly difficult problem due to the inherently dynamic and partial coverage of such systems.

The ultimate aim of this thesis is to show the feasibility of mobile sensor networks for monitoring air quality and their ability to achieve the goal of pushing our understanding of urban air pollution. To this end, we have taken a holistic approach, by studying the end-to-end system, and addressing each of the aforementioned challenges.

Using public transportation vehicles for mobility, we have developed and deployed a full-scale

Abstract

mobile sensor network for monitoring the air quality in the city of Lausanne, Switzerland. We have carefully considered all steps of the system design process, starting from the choice of targeted pollutants, sensor selection, node design, server architecture, and network operation and maintenance. We have opted for a modular sensor node architecture, which ensures a high degree of flexibility for future system augmentation or upgrades.

To ensure data quality, we investigated both the problem of mobility-caused distortion on slow-responding chemical sensor measurements, and that of sensor calibration. For addressing the former, we created a rigorous experimental set-up, inside a wind tunnel, in order to study this effect and the appropriate techniques for mitigating it. We are the first to study this problem in detail. We propose the use of deconvolution for recovering the underlying pollutant concentration signal, and demonstrate its effectiveness experimentally. Since the performance of this approach is limited by the signal-to-noise ratio of the measurements, we propose the use of an active sniffer to enhance the quality of the raw signal. The benefit of adding this active sampling system is demonstrated through both wind tunnel and outdoor experiments.

On the topic of sensor calibration, we propose two novel improvements to online rendezvous calibration methodology. The first one is a model-based approach, which considers the use of more sophisticated sensor models, which are more faithful to the complex behavior of chemical sensors. The second calibration enhancement builds upon our work on mitigating mobility-caused distortion, and proposes the use of a pre-processing step, in which the mobile measurement data is deconvolved. We illustrate the potential of these techniques for calibration enhancement by using simulated and real rendezvous data.

Finally, we study the problem of generating high-resolution maps based on mobile sensor network data. We address the problem of dynamic and partial coverage through the use of statistical modeling. We propose five modeling methods that use particulate matter data gathered by our network and a heterogeneous list of explanatory variables (e.g., proxy pollutants, meteorological data, land-use and traffic data). Three of these are very novel and make use of a virtual dependency network that relates street segments to each other. We evaluate and illustrate the ability of these techniques to generate high-resolution pollution maps, by using more than 44 million mobile particulate matter measurements.

Keywords: mobile wireless sensor networks, air quality monitoring, mobile chemical sensing, air sampling systems, chemical sensor calibration, high-resolution pollution mapping

Résumé

L A population urbaine augmente rapidement et de millions de décès sont attribués chaque année à la pollution de l'air. Dans ce contexte, le besoin urgent d'une compréhension plus approfondie de la qualité de l'air urbain est évident. La caractéristique locale des sources d'émissions urbaines, et la topographie spécifique des villes mènent à un paysage de concentration de polluants très hétérogène, dont les détails ne peuvent pas être captés par les stations de surveillance traditionnelles. Bien que très précis, ces systèmes sont grands, lourds et très coûteux, ce qui conduit à une distribution très diluée des mesures (par exemple, en Suisse seulement 16 stations sont disponibles pour l'ensemble du pays).

Les réseaux mobiles de capteurs sans fil ont le potentiel de permettre un changement de paradigme dans notre compréhension de la pollution de l'air en milieu urbain, à travers une augmentation significative de la résolution spatiale des mesures. Cependant, le chemin pour arriver à une information fiable et de haute qualité à partir de ce type de systèmes est entamé de défis. Ceux-ci commencent déjà à partir du niveau de la conception du système, étant donné que le développement de réseaux robustes de détection mobiles à grande échelle pour une surveillance urbaine continue est une tâche difficile en elle-même. Les limites de la technologie des capteurs existants est une autre source importante des problèmes difficiles. Les candidats typiques pour ce type d'applications, les capteurs chimiques souffrent d'une grande liste de problèmes qui rendent leur utilisation dans un scénario mobile non trivial. Ceux-ci comprennent : l'instabilité, la sensibilité croisée, de faibles rapports signal sur bruit, et une réponse dynamique lente. Ce dernier problème, en particulier, est un défi difficile lorsque l'on considère un scénario mobile, car il conduit à une distorsion significative des mesures. La question de la réalisation et le maintien du calibrage des capteurs chimiques, tout au long de la durée de vie de déploiement, est une autre question essentielle qui découle de leur instabilité typique. Enfin, le développement des techniques de modélisation appropriées qui nous permettrait de générer des cartes de pollution en haute définition sur la base des données de réseau de capteur mobile est un problème très difficile, en raison de la couverture par nature dynamique et partielle de ces systèmes.

Le but essentiel de cette thèse est de montrer la faisabilité des réseaux de capteurs mobiles pour la surveillance de la qualité de l'air et leur capacité à atteindre l'objectif d'augmenter

notre compréhension sur la pollution atmosphérique urbaine. À cette fin, nous avons adopté une approche holistique, en étudiant le système de bout en bout, et en abordant chacun des défis mentionnés ci-dessus.

En utilisant des véhicules de transport public en tant que source de mobilité, nous avons développé et déployé un réseau de capteurs mobile à grande échelle pour la surveillance de la qualité de l'air dans la ville de Lausanne, en Suisse. Nous avons examiné attentivement toutes les étapes du processus de conception du système, à partir du choix des polluants ciblés, au choix des capteurs, la conception de nœud, l'architecture du serveur, et le fonctionnement et la maintenance du réseau.

Pour aborder le problème de distorsion des mesures de capteurs chimiques causées par la mobilité, nous avons créé un montage expérimental rigoureux, à l'intérieur d'une soufflerie, afin d'étudier cet effet et les techniques appropriées pour l'atténuer. Nous sommes les premiers à étudier ce problème en détail. Nous proposons l'utilisation de la déconvolution pour récupérer le signal correct, et nous démontrons son efficacité par une approche experimentale. Puisque la performance de cette approche est limitée par le rapport signal-bruit des mesures, nous proposons l'utilisation d'un renifleur actif pour améliorer la qualité du signal brut. L'avantage de l'ajout de ce système est démontré par des expériences dans la soufflerie et dans le plein air.

Concernant la calibration des capteurs, nous proposons deux nouvelles améliorations à la méthodologie de calibration basée sur des rendez-vous. La première est une approche basée sur un modèle qui considère l'utilisation des modèles de capteurs plus sophistiqués, qui sont plus fidèles au comportement complexe des capteurs chimiques. La seconde amélioration de calibration se fonde sur notre travail sur l'atténuation des distorsions causées par la mobilité, et propose l'utilisation d'une étape de pré-traitement, dans lequel les données de mesure mobile sont déconvoluée. Nous illustrons le potentiel de ces techniques pour l'amélioration de calibration en utilisant des données de rendez-vous réelles et simulées.

Enfin, nous étudions le problème de la production des cartes à haute résolution à partir des données de réseaux de capteurs mobiles. Nous abordons le problème de la couverture dynamique et partielle en utilisant la modélisation statistique. Nous proposons cinq méthodes de modélisation qui utilisent des données de particules recueillies par notre réseau et une liste hétérogène des variables explicatives (par exemple, les polluants proxy, les données météorologiques, l'utilisation des terres et de données de trafic). Trois d'entre eux sont très nouvelles et font usage d'un réseau de dépendance virtuelle qui rapporte les segments de rue les uns aux autres. Nous illustrons la capacité de ces techniques pour générer des cartes de pollution en haute définition, en utilisant plus de 44 millions de mesures mobiles.

Mots-clés : réseaux mobiles de capteurs sans fil, surveillance de la qualité de l'air, détection mobile des substances chimiques, systèmes de prélèvement d'air, calibration des capteurs chimiques, cartographie de la pollution en haute résolution

Contents

Contents

Contents

Introduction Part I

1 Urban Air Pollution

I N March 2014, the World Health Organization released a statement in which its previous estimate on the number of premature deaths attributable to air pollution was more than doubled, reaching around 7 million annually [1]. This major increase was justified by both a better understanding of the diseases caused by air pollution, and, significantly, by improved estimations of human exposure to pollution through the use of new measurement technologies. That being said, our capacity to assess and predict air quality parameters on a scale relevant to human health remains quite limited. The established method for measuring air quality is the use of networks of static air pollution monitoring stations. These reference stations are equipped with expensive and highly accurate measurement systems, but are generally very sparse (e.g., the Swiss National Air Pollution Network - NABEL - has a total of 16 stations out of which only 6 are at urban or sub-urban sites).

At urban scale, factors like locality of emission sources, source strength (e.g., number of vehicles on a road), meteorology, natural and urban topography lead to highly heterogeneous concentration levels [2]. This issue makes the task of estimating human exposure based on traditional static stations very challenging, since using their measurements as surrogates for local concentration levels may significantly overestimate or underestimate personal exposure to pollutants.

1.1 Pollutant Types, Sources, and Distribution

A basic understanding of the targeted phenomenon is a prerequisite for a successful monitoring system design. For this reason, we discuss in this section the pollutants of concern in the urban environment, their basic characteristics, urban emission sources and typical distribution.

Air pollutants can be broadly classified as being either primary or secondary. Primary pollutants are defined as pollutants that are emitted directly from a source (e.g., vehicle exhaust, chimney stacks, etc.). Secondary pollutants are formed when other precursors substances

(e.g., primary pollutants) react in the atmosphere.

In the U.S., the Environmental Protection Agency (EPA) has identified six common air pollutants as "criteria pollutants" because of their impact on health and the environment [3]. They are carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), particulate matter (PM), and lead (Pb). These coincide broadly with the traditional air pollutants considered by the European Environment Agency (EEA) [4].

In our work we focus on this list of criteria pollutants. However, this is not an exhaustive urban air pollutant list. Other pollutants that we do not specifically consider are: mercury (Hg), cadmium (Cd), arsenic (As), and volatile organic compounds (VOCs) [5].

1.1.1 Carbon monoxide

CO is a primary pollutant that results from incomplete combustion of fossil fuels and biofuels. It is a colorless, odorless, and tasteless gas. While in the E.U. the introduction of Euro vehicle standards has lead to CO emissions from the transportation sector to fall second behind commercial, institutional and household fuel combustion [5], in the U.S., on-road traffic still accounts for 75% of all CO emissions in metropolitan areas [6].

CO is removed from the atmosphere mostly through reactions with hydroxil radicals (OH), to form CO_2 and hydrogen radicals that rapidly react with oxygen (O_2), forming HO_2 radicals. The typical lifetime of CO in urban environments is approximately 16 days, meaning that it is usually not destroyed inside the emission area, but is dispersed and mixed on a regional or continental scale [6].

CO levels can vary widely within a city depending on proximity to sources and topography [6]. Ambient CO typically exhibits a two-peak daily curve, which is driven by traffic flow dynamics [5], [6]. Typical ambient concentrations for CO range between 0-0.3 ppm [7], but personal exposure studies have shown that the highest CO exposure levels are registered inside vehicles, where CO concentrations can be 2 to 5 times higher than ambient values [8].

The largest percentage of time a person is exposed to CO is indoors. This means that high roadside CO concentration can represent an important health risk for people living or working in buildings that are adjacent to busy roads, since all of the ambient CO infiltrates indoors [6].

Carbon monoxide exposure has been shown to cause an increase in cardiovascular and respiratory morbidity when considering short-term exposures, and also to have negative effects on the central nervous system for both short- and long-term exposures [6].

1.1.2 Nitrogen dioxide

NO_2 is both a primary and a secondary reactive pollutant. It is a reddish-brown toxic gas (see Figure 1.1) and has a characteristic sharp, biting odor. Since they are emitted together and can

Figure 1.1 – Nitrogen dioxide (NO_2) has a reddish-brown color. The intensity of the color varies with temperature, as NO_2 converts to the colorless gas dinitrogen tetroxide (N_2O_4) at low temperatures, and converts back to NO_2 at higher temperatures. The bottles in this photograph contain equal amounts of an NO_2/N_2O_4 mixture at different temperatures. (Image source: Public Domain)

interconvert rapidly, NO_2 is often grouped with nitrogen monoxide (NO) under the collective term NO_x. The conversion from NO to NO_2 has a timescale of approximately one minute [9].

The main source of NO_x is the transportation sector [5], [9]. Most direct NO_x emissions are in the form of NO, with a small percentage (about 5% to 10%) being NO_2. Diesel powered vehicles are, however, an important exception, with NO_2 representing as much as 60% of NO_x emissions, for vehicles that use new oxidation catalyst technologies [10].

Beyond emission sources, ambient NO_2 levels are influenced by its high reactivity, meteorology, and surface deposition. NO_2 is one of the precursors to tropospheric O_3. It reacts in the air and on surfaces, forming various oxidation products, like peroxyacyl nitrates (PANs), a group of powerful eye and respiratory irritants, or nitric acid (HNO_3), a highly corrosive acid. The timescale for these reactions ranges from a few hours during summer to approximately one day in winter. Eventually, oxidized nitrogen compounds are removed from the atmosphere by surface deposition [9].

In terms of spatio-temporal variability, NO_2 behaves mostly as a primary pollutant being linked to mobile pollution sources (i.e. vehicles). As a consequence, NO_2 exhibits significant small-scale horizontal and vertical gradients when moving away from the emission sources. Temporally, NO_2 concentrations are driven by daily traffic dynamics, and, on a seasonal scale, by meteorological processes, which lead to typical higher levels in winter when the mixing layer height is lower. The typical concentration range for NO_2 is 0-50 ppb [7].

Short-term exposure to NO_2 has been linked to an increase of respiratory morbidity (e.g., reduced lung function, inflammation, airway hyperresponsiveness) [9].

1.1.3 Sulfur dioxide

SO_2 is a primary pollutant resulting from combustion of sulfur containing fuels. It is a colorless, toxic gas with a pungent, irritating smell. The main source of SO_2 is the energy sector,

accounting for more than 60% of emissions, followed by the industrial sector, with more than 24% [5], [11]. A major natural source for SO_2 emissions are volcanic eruptions.

SO_2 is a precursor of acid rain, going through a series of reactions in the atmosphere to form sulfuric acid (H_2SO_4), which attaches to aerosol particles and cloud droplets. The typical lifetime of atmospheric SO_2 ranges from under one day to four days, depending on altitude and meteorological conditions [11].

Since SO_2 emissions are mainly caused by power generating stations and volcanoes, the relevance of including SO_2 on a list of monitored urban air parameters depends on the existence of such sources in the considered environment. Moreover, since these sources are static and the emission parameters of power stations (e.g., stack heights, emission rates) are monitored and usually regulated, SO_2 levels can be estimated spatio-temporally with a significantly lower uncertainty, based on traditional monitoring techniques.

SO_2 exposure can aggravate pre-existing respiratory conditions, and, while asthmatics are most at risk, high-levels can cause respiratory symptoms for people without a history of pulmonary disease [11].

1.1.4 Ozone

O_3 is a molecular form of oxygen composed of three oxygen atoms, much less stable than the more common O_2. It is a pale blue gas with a pungent smell. While in the stratosphere O_3 occurs naturally and serves the beneficial role of absorbing harmful solar radiations, in the troposphere, it is a powerful oxidizer, harmful to living organisms and materials. In urban environments, O_3 is a secondary pollutant formed by complex reactions involving precursor pollutants (e.g., NO_x, CO, and VOCs) in the presence of sunlight. A key element of tropospheric O_3 generation is the photolysis of NO_2:

$$NO_2 + h\nu \rightarrow NO + O(^3P) \tag{1.1}$$
$$O(^3P) + O_2 \rightarrow O_3 \tag{1.2}$$

In the presence of light, NO_2 forms NO and releases an oxygen atom, denoted as $O(^3P)$ in Equation (1.1). The atomic oxygen subsequently reacts with molecular oxygen, to produce O_3 (see Equation (1.2)). This is a reversible reaction, as O_3 can react with NO to produce NO_2 and O_2:

$$O_3 + NO \rightarrow NO_2 + O_2 \tag{1.3}$$
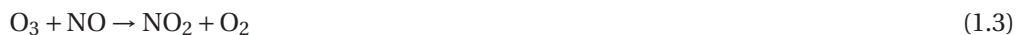
For a given solar radiation condition, Equations (1.1) to (1.3) form an equilibrium called the photo-stationary state.

Ground level O_3 varies spatially because of reactions with other chemical species. For example near busy roads ozone levels decrease due to reaction with vehicle emitted NO [9], [12]. Because of its relationship with sunlight, ozone also varies temporally at different scales, showing daily cycles (higher levels during daytime vs. nighttime), and also significant seasonal variation. The typical concentration range for O_3 is 0-150 ppb [7].

Exposure to O_3 has been proven to cause a variety of respiratory problems including decreased lung function, lung inflammation, and airway hyperresponsiveness [12].

1.1.5 Particulate matter

Atmospheric PM refers to microscopic solid or liquid matter suspended in the atmosphere. Unlike gas-phase pollutants, PM varies in size, shape and chemical composition. It is both a primary and secondary pollutant. The main sources of primary PM emissions in cities are vehicle exhausts, road dust resuspension, and domestic heating [5]. The main gaseous precursors to secondary PM are SO_2, NO_x, VOCs, and ammonia (NH_3) [5], [13].

Multiple metrics are used to quantify particulate matter in a given volume, the most common being particle mass, particle count and surface area. PM is usually classified by size range, with particles between $2.5\,\mu m$ and $10\,\mu m$ (PM_{10}) representing coarse particles, and particles smaller than $2.5\,\mu m$ ($PM_{2.5}$) being usually labeled as fine particles. Particles that are smaller than $0.1\,\mu m$ are called ultrafine particles (UFPs), and are of particular interest for health. While both PM_{10} and $PM_{2.5}$ can cause important health problems, fine particles can travel deeper into the lungs and can have worse effects than the more coarse particles.

PM negative health effects include triggering and aggravation of pulmonary and cardiovascular condition, and increased mortality [13]. Ultrafine particles (UFPs), defined as particulates with a diameter smaller than $0.1\,\mu m$, are of particular health interest with new studies pointing to a potentially higher toxicity due to their large surface-to-size ratio [14], [15], and also their ability to cross the cellular membrane [16].

PM can exhibit important spatial variability due to locality of emission sources, with near road environments presenting high concentration gradients, in particular for UFPs [17]. In the temporal dimension, urban particulate matter concentrations are influenced by anthropogenic (e.g., traffic dynamics, domestic heating patterns) and meteorological processes.

PM is removed from the atmosphere through dry or wet deposition on surfaces. Dry-deposition happens mostly through surface impact and gravitation, and it affects mostly the coarse range of PM. Wet deposition implies either an initial incorporation of particles into cloud droplets, followed by precipitation, or the scavenging of particles by rain drops or snow flakes during their fall [13].

1.1.6   Lead

Historically, the major source of lead pollution has been from fuels used for on-road vehicles and industry. As a result of regulatory efforts to ban leaded gasoline additives, vehicle Pb emissions have been mostly eliminated in industrialized countries. Today, the main sources of lead emissions are from piston-engine aircraft and the industry sector (e.g., metalworking plants, cement factories).

In the atmosphere, Pb is associated to PM in both the coarse and fine ranges. As other PM, it gets removed from the atmosphere through wet or dry deposition, with coarse lead PM being deposited close to the industrial sources which have produced them, while fine lead PM can be transported over very long distances [18].

Lead is a toxic heavy metal that can damage the developing nervous system of infants and causes cardiovascular and renal disease in adults.

1.2   Human Exposure and Significance of High-Resolution Data

Human exposure to an airborne pollutant is defined as the contact between the human and the pollutant over a certain period, called the exposure period. It can be expressed mathematically in the general form:

$$E = \int_{t_0}^{t_f} C(t) dt \tag{1.4}$$

where E represents the exposure over an exposure period that ranges from t_0 to t_f, and $C(t)$ is the concentration of the considered pollutant at time t.

We consider only the ambient component of human exposure to air pollution, as a consideration of the non-ambient component (e.g., from cooking, smoking, etc.) is beyond the scope of our work. Using the same formulation as in [6], the human exposure to an ambient pollutant, E_a, can be expressed in terms of fractions of time spent in different micro-environments as:

$$E_a = \sum_o f_o C_o + \sum_{o,i} f_i F_{inf,i} C_{o,i} \tag{1.5}$$

where:

- f_o and f_i are the fractions of the total exposure time spent in outdoor and, respectively, indoor micro-environments.
- C_o is the average pollutant concentration in outdoor micro-environment o.
- $C_{o,i}$ is the average pollutant concentration in the outdoor micro-environment o adjacent to indoor micro-environment i.
- $F_{inf,i}$ is the infiltration factor for indoor environment i, which is a function of the building's air exchange characteristics.

Because of the lack of high spatial-resolution measurement data, epidemiological studies often use ambient concentration measurements C_a from static monitoring stations for all outdoor micro-environments [6], [9], [12]. This leads to the following simplified form for the ambient exposure:

$$E_a = (f_o + \sum_i f_i F_{inf,i})C_a \tag{1.6}$$

The simplifying assumption behind Equation (1.6) can, however, be inappropriate when a high spatial heterogeneity is present in the pollutant field because of variability of emission sources, meteorology and topography.

This issue can be addressed through the use of spatial estimation models. Spatial estimation models try to augment the spatial resolution of the typically sparse static monitoring stations by constructing a complete pollution map of the area of interest, using additional sources of information (e.g., land-use designation, traffic counts, meteorologic data, etc.). This type of models includes *land-use regression models* [19], but also the more complex and computationally expensive *dispersion models* [20], [21].

Although to different degrees, these modeling techniques suffer from uncertainties and resolution limitations of set-up parameters (e.g., emission inventories, traffic volume profiles) and would greatly benefit from more distributed measurements for both model derivation and validation. Furthermore, these models cannot adapt without manual intervention to real-world changes respective to their set-up parameters (e.g., topography or traffic flow modifications), and cannot capture particular pollution events (e.g., construction sites, fires, etc.).

Summary

In this chapter we provided an overview on the topic of urban air pollution. We presented a list of urban air pollutants monitored by both the United States EPA and the European Union EEA because of their impact on human health. These are: carbon monoxide, nitrogen dioxide, sulfur dioxide, ozone, particulate matter, and lead. We presented their basic properties, emission sources, typical urban distribution, as well as their proven impact on human health. Finally, we discussed the issue of accurately estimating human exposure to pollutants and presented arguments for the opportunity of increasing the spatial resolution of air quality measurements in cities.

2 Air Quality Monitoring

A IR quality monitoring is performed through the use of specialized measurement equipment for investigating air pollution and its impact on human health and the natural environment. Air monitoring equipment is operated by governmental agencies, researchers, or private citizens.

In this chapter we present an overview of air quality monitoring systems, including both established systems and developing technology. We are interested only in systems that are relevant for the urban environment, and as such do not consider satellite-based remote sensing technology (e.g., the Ozone Monitoring Instrument on board the NASA Aura satellite), which have a low spatio-temporal resolution and are appropriate for long-term pollutant trend analysis at regional and continental scales.

2.1   Established Measurement Systems

The traditional approach for long-term monitoring of air quality is the use of national or regional *networks of static monitoring stations* which are operated by governmental agencies. These stations are large and equipped with expensive and highly accurate measurement systems for air pollutants (e.g., absorption spectrophotometers, mass spectrometers, etc.). They also typically monitor relevant meteorological parameters (e.g., temperature, precipitation, wind, etc.). The high cost of these stations means that they form very sparse networks.

An example of such a system is the National Air Pollution Monitoring Network (NABEL[1]) in Switzerland (see Figure 2.1). It consists of 16 monitoring stations located throughout the country at sites that are specially selected in order to cover different classes of locations: urban with traffic, urban residential, suburban, rural, etc. NABEL's mission is to monitor air pollution on a national level and gauge the impact of environmental policies. While the temporal resolution of these stations is high, their sparseness means that they cannot provide a detailed image of intra-urban air pollution.

---

[1] http://www.bafu.admin.ch/luft/00612/00625/index.html
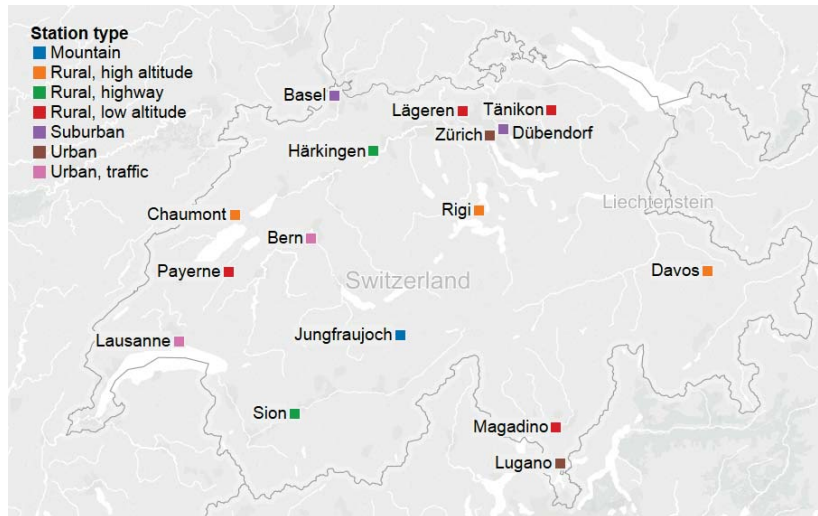
Figure 2.1 – The Swiss NABEL network is comprised of 16 monitoring stations, out of which only 6 are placed in an urban or suburban location.

Personal monitoring is an established method for gathering air pollution data for human exposure research. It is typically used in epidemiological panel studies [22], [23]. These imply giving some portable equipment, usually placed inside a backpack, to individuals who carry it around for a predefined period of time (e.g., 24 hours), at regular time intervals. The backpack is then returned for data transfer and processing.

Diffusive passive samplers are often used, particularly for measuring NO_2 and O_3 exposure (e.g., Palmes tubes [24], Yanagisawa badges [25], etc.). These devices are very cheap and semi-disposable. They work by accumulating pollutants over time and do not provide any information on short-term variations. In order for their exposure levels to be extracted, they need to be processed and analyzed in a lab.

For other pollutant species such as CO and PM, portable loggers are available but their cost is usually high. This means that even if the overhead of processing is lower than for passive samplers, the equipment cost significantly limits the number of devices available for panel members, and consequently the panel size, leading to a question of representativeness.

While providing important information on pollutant exposure on the individual level and the relationship between ambient and indoor pollutant concentrations (see parameter $F_{inf,i}$ in Equation (1.5)), studies employing personal monitoring have important scaling limitations which arise from their specific technical and economical constraints, and cannot be considered an alternative to long-term air quality monitoring infrastructure.

Figure 2.2 – General representation of a WSN. All WSNs are composed of sensor nodes spread over a physical domain, that c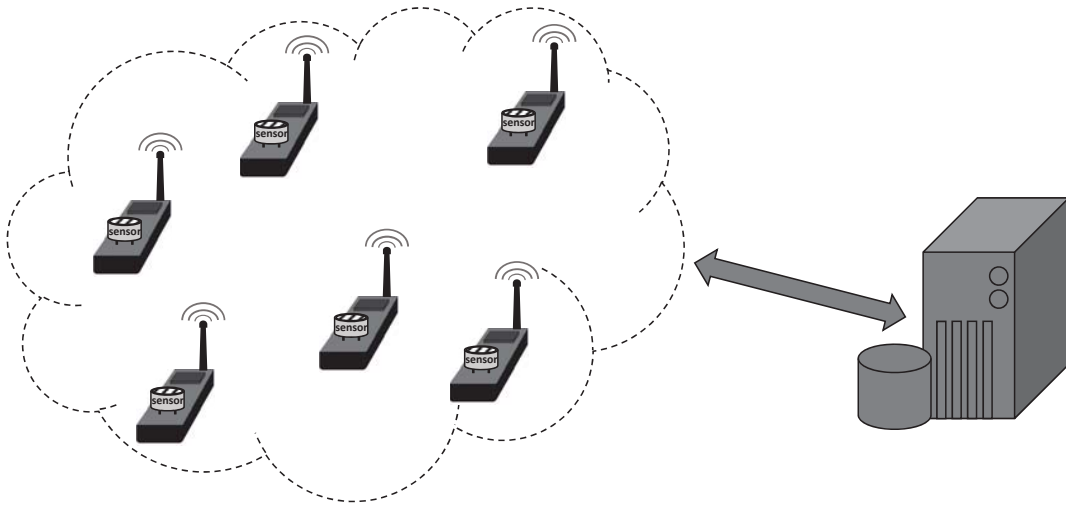an take measurements and send them over a wireless communication link (directly or through multiple hops) to a back-end server for storage and processing.

2.2 Wireless Sensor Networks for Air Quality Monitoring

The Wireless Sensor Network (WSN) paradigm proposes the use of relatively cheap sensing nodes (i.e. devices with embedded computation, sensing and communication) in large numbers over a physical area in order to obtain high resolution data of a targeted phenomenon (see Figure 2.2).

Although not the case for all applications, moving from monolithic monitoring systems to highly distributed systems like WSNs typically implies a trade-off in measurement quality, as the sensors that are small enough and cheap enough to be used in this context suffer from various issues including high sensor noise, poor sensitivity, drift, etc. These issues are certainly present in applications targeting air quality monitoring with WSNs, making the tasks of system design and measurement processing very challenging. Nevertheless, WSNs do hold a great potential for complementing more traditional air monitoring systems and delivering the type of spatio-temporal density of measurements needed to advance our understanding of human exposure to airborne pollutants.

Traditionally, most of the research in the WSN field has focused on static sensor networks, but recent years have seen a rapidly growing interest in systems that include mobility. Mobility is certainly preferred for air monitoring applications, since the deployment and maintenance cost for covering an entire city with an exclusively static deployment, with the resolution of a few tens of meters, would still be prohibitively high. As such, the number of projects that do not consider mobility is low and usually restricted to smaller area studies, most projects considering either a mix of static and mobile nodes, or exclusively mobile deployments. One example of completely static network is the one used for Sensor Networks for Air Quality at

Heathrow Airport (SNAQ-Heathrow) project [26], which targeted the study of air quality over the airport area using a dense network of fixed sensor nodes.

When endowing a sensing system with mobility, there are two main classes to choose from: controlled or uncontrolled mobility. Controlled mobility is the object of research of mobile robotics which has produced a wide array of different platforms used to control the trajectories of the sensing nodes in water [27], on land [28] or airborne [29].

While controlled mobility permits the freedom to choose desirable trajectories and formations, this comes at a high energetic price, since, besides the energy consumed for measuring and transmitting data (which already represents an important traditional problem for WSNs), one also needs to use a significant amount of energy to keep the system mobile. In order to have long-time deployments of large robotic networks for real-world applications, the hard problem of energetic autonomy needs to be addressed.

By using uncontrolled parasitic mobility, this issue is avoided by anchoring sensor nodes to already existing mobility sources. In an urban scenario, these range from public transportation and private vehicles to cyclists and pedestrians.

However, the use of any kind of mobility for an air quality monitoring application gives rise to a number of research challenges. The main difficulty arises from the fact that the chemical sensors that are compact and inexpensive enough to be used in a sensor network solution have typically significantly slow dynamics relative to the motion of the available urban mobility vectors.

2.2.1 Networks of portable sensor nodes

Started in 2006, the Mobile Environmental Sensing System Across a Grid Environment (MES-SAGE) project [30] was one of the first European projects to study the use of dense networks of low-cost sensors for monitoring air quality. Both static and portable, hand-held sensing platforms were used [31], [32]. It was a precursor of the SNAQ-Heathrow project.

CitiSense [33] is a more recent project aiming at providing a citizen infrastructure for monitoring pollution and environmental conditions that users are exposed to. The idea is to collect data on pollutants, using mobile phones and small sensors carried by users. It aims at designing a complete cyber-infrastructure that addresses issues of power management, communication, data security and privacy. Pilot studies planed in this project focus on a range of environmental issues of societal concern, such as research on the association of environmental exposure with human health, noise and development of public spaces, and monitoring indoor air quality at schools. On the other hand, in terms of pollutant sensing, this project uses very low-cost and low-quality sensing platforms.

SmartSantander [34] was another project aiming at developing cyber-infrastructures for enabling community sensing focused in smart cities. One of the goals of this project was to

monitor the air quality by the participants. However, the project does not go into the details of the sensing technologies.

EveryAware [35] was a project whose goal was to create new technological platforms combining sensing technologies, networking applications and data-processing tools for community-based air pollution monitoring. This project has systematically investigated the calibration and performance of many commercial gas sensors for indoor/outdoor stationary/mobile platforms, focusing on estimating the concentration of Black Carbon (BC), a component of $PM_{2.5}$, based on other measured pollutants (e.g., CO and NO_2). Another focus of the project has been on developing user interfaces and relevant applications for smart phones.

The AIR (Area's Immediate Reading) project [36] also investigated participatory air quality sensing by giving about 10 sensor boxes to the public to explore the urban environments. The sensor box measured CO, NO_x and O_3. This project also did not go into the details of sensing technologies, although accurate and high tech sensing nodes were used.

Common Sense [37] developed low-cost, hand-held sensors to be used by participants and studied mechanisms for understanding and using input data from various users (e.g., community members, scientists, etc.). The project explored the process of knowledge production through several dozen interviews with the users [38]. Unlike other systems, it derived design principles and a framework for describing data through these interviews, aiming at user-centered design of a system for community analysis of air quality data.

For the AirCloud system [39] two types of $PM_{2.5}$ monitors were developed: the larger static AQM, which has a direct data up-link through Ethernet or GPRS, and the portable miniAQM, which is designed to interface through a smartphone via Bluetooth. On the server side, an air-quality analysis engine was developed, which learns sensor calibration models using Artificial Neural Networks (ANN), and estimates the $PM_{2.5}$ field using Gaussian Processes.

2.2.2 Bicycle-mounted sensor networks

Bicycles have also been used as a vector of mobility for air quality monitoring. One example is the Copenhagen Wheel project [40], which anchored environmental sensors on the rear wheel of a bicycle, providing information that could add value to the rider's experience, and also providing measurements to city planners (including CO, NO_x, noise, ambient temperature and relative humidity).

The Aeroflex sensor bike [41], developed by the Flemish Institute for Technological Research (VITO), is a regular bicycle equipped with instruments to measure local air quality. The bike measures UFP count (using the TSI P-Trak particle counter), PM_{10}, $PM_{2.5}$ and PM_1 (using the Grimm Technologies 1.108 aerosol spectrometer), BC (using the AethLabs microAeth AE51 aerosol monitor), and CO (using the Alphasense CO-BF electrochemical cell sensor) at a high temporal resolution. The measurements are linked to their geographical location and time using a GPS module and Internet time. Furthermore, the Aeroflex is equipped with automated

data transmission, data pre-processing and data visualization systems.

2.2.3   Vehicular sensor networks

The Mobile Air Quality Monitoring Network (MAQUMON) project carried out by the Networked Embedded Systems Lab at ISIS, Vanderbilt University, built a prototype system comprised of sensor nodes mounted on cars. The sensor node consists of a microcontroller, an on-board GPS unit, and a set of gas sensors measuring O_3, CO, and NO_2 concentrations. The node is Bluetooth enabled, so it can connect to a smart phone or laptop to upload the measurements [42].

Haze Watch [43] was another project in which mobile phones were interfaced (through Bluethooth) with external pollution sensors attached to cars, in order to measure the concentration of CO, O_3, SO_2, and NO_2 in the air. The time-stamped and geo-tagged measurements are then uploaded to a server to build maps, which aggregate the readings of all participants and are accessible by the public.

In [44], the concept of a mobile air monitoring system is presented using two types of sensing platforms: a custom-built Mobile Sensing Box (MSB) destined for public transportation vehicles, and a Personal Sensing Device (PSD) targeting private vehicles drivers, based on the Variable Inc. NODE system. The presented results for both platforms are, however, based on experiments using a single car.

The Citi-Sense-MOB [45] is an European project focused on the development of an infrastructure to continuously monitor environmental data at street level using sensors mounted on mobile platforms, such as buses and bicycles. This project also develops a user interface, using both web services and mobile phone applications.

Using public transportation as the source of mobility has obvious advantages. First, by design, public transportation networks should ensure good coverage of a city in both space and time. Second, using this type of mobility simplifies the maintenance of the system which can be done at a single or a few specific points (i.e. the depot(s)). Considering that the most significant source of urban pollution is vehicle exhaust, measurements taken within the traffic flow should provide an advantage in estimating the overall pollution field. Finally, the possibility to use the vehicle's energy reserve is certainly very attractive, since this can power both the measurement and communication tasks, providing full energetic autonomy. For all these reasons, in our work we have considered public transportation-anchored nodes as the back-bone of our air quality monitoring WSN system.

Summary

In this chapter, we introduced the state of the art in air quality monitoring systems. We presented established systems for gathering air pollution data. These include government-operated networks of static stations for long-term pollution monitoring, and personal monitoring devices typically used in human exposure studies by health researchers. We provided an extensive review of previous and ongoing work in the field of wireless sensor networks for air quality monitoring. In this application domain, because of economical and coverage reasoning, mobility is typically used, including networks of portable sensor nodes, bicycle-mounted networks and vehicular sensor networks. Finally, we justified our choice for a public transportation-based air quality monitoring network, through the many advantages provided by this mobility vector (e.g., for urban coverage, maintenance, proximity to emission sources, and power autonomy).

3 Scope of this Thesis

ENABLING high spatio-temporal resolution monitoring of urban air quality through Mobile Wireless Sensor Networks is the focus of this thesis. For this purpose, we adopted a holistic approach going progressively through all the steps of a complete system design. This included the development and deployment of a full-scale mobile WSN using public transportation vehicles, the study of techniques for mitigating undesirable effects of mobility on measurements, sensor calibration, and measurement-driven high-resolution air pollution mapping.

3.1 Objectives and Outline

The outline of the thesis is represented in Figure 3.1. It has an incremental linear logical structure, an appropriate approach to each preceding module being a prerequisite for the successful handling of the next.

In the first module (**Part II** of this manuscript), we go through all the necessary steps for a successful *design of a mobile WSN for air quality monitoring*, from selecting the targeted pollutants, choosing the appropriate sensors, designing the sensor node and server architecture, to the network deployment and operation. A relevant paper for this part, currently in preparation for journal submission, is:

- ■ **A. Arfire**, A. Marjovi, A. Bahr, E. Droz, J. Eberle, J.-P. Calbimonte, K. Aberer, and A. Martinoli, "Sensing the Air We Breathe: Designing a Mobile Air Quality Monitoring
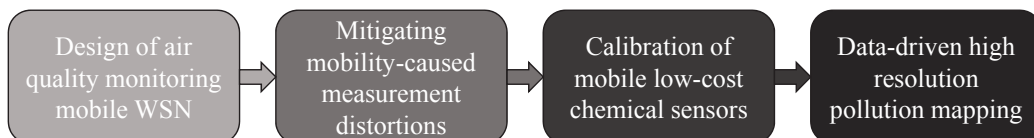

Figure 3.1 – Thesis outline.

Network", 2016, Working paper.

The second module (**Part III** of this manuscript) investigates an essential challenge for the success of mobile WSN applications that employ low-cost chemical sensors: *mobility-caused measurement distortion.* Most commercially available chemical sensors have a long response time (i.e., slow dynamic response to stimuli), ranging from a few tens of seconds to more than a minute. Thus, anchoring these sensors to a fast moving platform leads to significant signal distortion respective to the underlying concentration field. We propose and evaluate different methods of mitigating this effect. Relevant papers for this part are:

- **A. Arfire**, A. Marjovi, and A. Martinoli, "Enhancing Measurement Quality through Active Sampling in Mobile Air Quality Monitoring Sensor Network", in *Proceedings of the 2016 IEEE International Conference on Advanced Intelligent Mechatronics*, 2016.
- **A. Arfire**, A. Marjovi, and A. Martinoli, "Mitigating Slow Dynamics of Low-Cost Chemical Sensors for Mobile Air Quality Monitoring Sensor Networks", in *Proceedings of the International Conference on Embedded Wireless Systems and Networks*, 2016, pp. 159–167.

The third part (**Part IV** of this manuscript) addresses a core problem of sensor network applications that employ low-cost chemical sensors, that of ensuring *sensor calibration* throughout the duration of the deployment. Chemical sensors have generally a poor selectivity, being cross-sensitive to other chemical species than the one targeted, but also to environmental parameter variations (e.g, in temperature, humidity, pressure, etc.). Moreover, their input-output characteristic is not stable in time, but drifts because of sensor aging. In the context of long-term large-scale mobile deployments, manual periodic re-calibration cycles are impractical, and the aforementioned issues need to be handled through automatic on-line calibration algorithms. A relevant publication for this part is:

- **A. Arfire**, A. Marjovi, and A. Martinoli, "Model-based Rendezvous Calibration of Mobile Sensor Networks for Monitoring Air Quality", in *Proceedings of the 2015 IEEE SENSORS Conference*, 2015, pp. 366–369.

The final part (**Part V** of this manuscript) focuses on the end-goal of our mobile air quality monitoring networks: providing *high-resolution urban air pollution maps*. We propose the use of different data- and measurement-driven statistical modeling algorithms for estimating road-level pollution concentrations. Relevant papers for this part are:

- A. Marjovi, **A. Arfire**, and A. Martinoli, "Leveraging Land-use Data and Machine Learning for Extending Urban Air Pollution Maps Beyond the Coverage of a Mobile Sensor Network", submitted to the 14th ACM Conference on Embedded Networked Sensor Systems, 2016.
- A. Marjovi, **A. Arfire**, and A. Martinoli, "High Resolution Air Pollution Maps in Urban Environments using Mobile Sensor Networks", in *Proceedings of the 11th International Conference on Distributed Computing in Sensor Systems*, 2015, pp. 11–20.

3.2 Contributions

The contributions of our work are grouped into four main parts, corresponding to the structure of our research outlined in Figure 3.1:

1. **Designing an air quality monitoring mobile WSN**
   Our mobile air quality monitoring network in Lausanne, Switzerland, is the result of more than three years of intensive research, development and prototype testing. It has been in full operation for more than two years. Through this process we have gained a unique first-hand experience and an insight into the potential and current limitations of this type of monitoring system. We have designed the system with two main objectives in mind: to bridge the distance between traditional methods of air quality measurement, and to constitute a flexible research platform for exploring mobile air measurement system designs. The latter goal is particularly unique to our work, as, to the best of our knowledge, no other group has targeted the development of a platform that would enable the future research of air sampling system designs. We consider this research direction essential for achieving reliable and accurate mobile air quality monitoring.

2. **Mitigating mobility-caused measurement distortions**
   Leveraging a wind tunnel facility, we are the first to study the impact of mobility on chemical sensing accuracy through a rigorous experimental framework. We showed that the effect of mobility is significant and that it has to be taken into account when working with chemical sensors for mobile air quality monitoring. We proposed the use of a deconvolution technique for reducing the distortion of the measured signals. We also studied the opportunity of using active sniffers for enhancing the quality of chemical sensor measurements. We considered both fan- and pump-based designs and studied their ability to improve the measured signal in terms of signal-to-noise ratio and response time. For the best performing sniffer, we went a step forward, and showed that its performance is maintained when moving outside of the wind tunnel, in a real-world comparative experiment using an electric car as mobility source.

3. **Calibration of mobile low-cost chemical sensors**
   We are the first to study the opportunity of using faithful sensor models in the context of online rendezvous calibration of mobile air quality monitoring networks. Studying the long-term performance of the electrochemical sensors used in our deployment for monitoring carbon monoxide, we derived models with increasing levels of complexity. These included models that explicitly take into account temporal drift and temperature effects. We proposed a model-based sliding-window rendezvous calibration algorithm and evaluated its performance when varying the complexity of the assumed sensor models and the size of the calibration window, using both simulated and real rendezvous data. Building upon our work on mitigating the effects of mobility on chemical sensor measurements, we also proposed a novel mobility-aware approach to rendezvous calibration. We evaluated the performance of this technique and showed its potential for improving calibration accuracy.

4. **Data-driven high-resolution pollution mapping**

   Five modeling methods using a real-world large scale mobile sensor network were proposed to generate high spatio-temporal resolution PM maps for an urban environment. The models can deal with dynamic coverage of the mobile sensor network. The first method was a conventional log-linear regression model based on nine meteorological and proxy pollutant explanatory variables. For the second method, a network-based log-linear regression model, we proposed creating a virtual network based on the dependencies of PM values in which each street segment is considered as one node of the graph and each edge represents correlations between two nodes. The first two models are able to estimate pollution levels only for the streets within the coverage of our mobile sensor network. To overcome this limitation we proposed to also use land-use and traffic count data as explanatory variables. The third, forth and fifth models we presented all make use of this additional sources of information, and the maps generated by them are able to go beyond the coverage of our sensor network. The third method is a direct extension of the basic log-linear regression model, by including also land-use and traffic data. The fourth model is a modified network-based log-linear regression model, with the dependency network being built based on land-use data. The fifth model we proposed uses Deep Neural Networks that allows the model to learn non-linear and complex dependencies between the model variables. The second, fourth and fifth modeling techniques have a high degree of novelty. We evaluated and discussed the performance of all the proposed techniques using multiple statistical metrics.

Addressing this diverse list of topics implied a collaborative endeavor, and although I was directly involved in all of the work presented in this thesis, I would like to give clear credit to my close collaborators. Ali Marjovi has been my closest research colleague, and has also been involved to varying degrees in all the research topics covered in this work. In particular, I need to credit him with assuming the lead on the design and implementation of statistical pollution modeling techniques.

Alexander Bahr and Emmanuel Droz contributed significantly in the development of the mobile sensor network used in this thesis. Credit goes to both of them for the design of multiple electronic boards included in our platform (e.g., GSM modem interface, sensor interfaces, etc.). Additional credit goes to Emmanuel Droz for the mechanical design of protective inlets for our O_3 and PM sensors, and for building a baseline code framework for the microcontroller used in our sensor nodes. This simplified the task of developing the node firmware.

Finally, I would like to give credit to Julien Eberle for leading the development effort for the server-side system architecture.

Summary

In this chapter, we introduced the outline and scope of this thesis. We provided an overview of the four modules of our work, including references to our relevant publications. The contributions of our work are associated with the four main parts of the thesis: mobile air quality monitoring WSN design, the mitigation of mobility-caused measurement distortions, chemical sensor calibration, and high-resolution urban air pollution mapping.

Designing a Mobile Air Quality Monitoring Network

Part II

4 Introduction

E MPOWERED by progress in sensor technology, wireless sensor networks hold the potential to constitute a real game changer for our understanding of urban air pollution and its impact on human health, through a significant augmentation of spatial resolution in measurement. However, the road to this new age of urban air monitoring is full of scientific challenges and technical pitfalls. In this part we share our first-hand experience in developing, deploying, operating and exploiting a mobile air monitoring network in urban settings. Using sensor nodes anchored to public transportation vehicles, our ongoing deployment has been gathering data on a significant collection of health-relevant pollutants for more than two years (33 months up to date).

In particular, we present the essential steps of developing this type of system including choosing the targeted pollutants, sensor selection, system design, deployment, and maintenance. Furthermore, we give an insight into what to expect in terms of system performance by analyzing network coverage and the ability of the system to capture spatio-temporal variability.

In this chapter, we start by discussing what are the design objectives for a mobile WSN for monitoring urban air quality. We then explain the rationale behind our choice of pollutants to monitor.

4.1 Design Objectives

A successful mobile WSN needs to be able to balance well a relatively wide set of design objectives. In this section, we present a list of the ones we consider to be the most significant, and discuss our approach to each of them. Our list includes:

1. **Measurement quality**
   Ensuring a certain level of measurement quality is obviously a critical aspect for any monitoring application. However, in the case of many existing platforms, particularly for

participatory monitoring (e.g., Air Quality Egg [1], Data Canvas [2]), this aspect has received less attention compared to providing platform accessibility. While compromises can be made between sensor cost and data quality when it comes to characteristics like stability, measurement noise, response time, or cross-sensitivity, this is not the case for sensor sensitivity. Ensuring that the limit of detection (LOD) of the selected sensors is not above the ambient concentration range is a critical condition.

2. **Coverage**

   An important performance parameter for a mobile WSN is its degree of spatio-temporal coverage. For a given number of sensor nodes, the choice of mobility platform is key to maximizing network coverage. By design, public transportation provides a high spatial coverage of urban environments. Furthermore, public transportation vehicles are in continuous operation for up to 20 hours per day, with a higher number of vehicles being allocated during rush hours, when pollution field dynamics are higher. Considering that during the night traffic emissions are much lower, the relatively short period when no buses are in operation can be complemented by a reduced set of well placed static stations.

3. **Robustness**

   Monitoring air quality is a long-term endeavor. The design of an outdoor WSN that is required to operate for long periods of time (in our case, multiple years) needs to take into account the large variations of meteorological parameters (e.g., temperature, humidity, precipitation) that the systems will have to withstand. In the case of air quality measurements, a careful balance needs to be struck between sensor exposure to the environment and overall system protection from dirt and water infiltration. Ensuring this protection is generally more difficult for mobile systems compared to static ones. Furthermore, a mobile platform needs to resist to important mechanical strains due to vehicle acceleration and vibration.

4. **Energetic autonomy**

   Although less critical than robustness, energetic autonomy is also important for long-term deployments of WSNs, as the use of replaceable batteries would complicate and increase the cost of network maintenance. Preserving energy is an important research topic in the field of WSNs with a lot of focus being placed on energy scavenging, low-power electronics, and power-saving communication protocols. When considering vehicular sensor networks, however, this issue is eliminated by the possibility to tap into the vehicle's power supply, which can usually provide enough power for all of the sensor node's functions.

5. **Maintainability**

   The ease to perform network maintenance has an important impact on deployment reliability, running cost and lifetime. It needs to be taken into account in the design process for both the sensor node and the back-end server. When considering a mobile field deployment with limited access for maintenance operations, the choice of modular

---

[1]http://airqualityegg.com
[2]http://datacanvas.org

hardware platform design is usually a good idea. This helps reduce the time needed for maintenance interventions as it permits the replacement of only the malfunctioning module. On the back-end server side, the implementation of automatic fault detection and diagnosis services can greatly help reduce the maintenance effort.

6. **Flexibility**

   The field of mobile WSNs for air quality monitoring is currently highly dynamic. This is due to both a high interest in the subject within the research community, and also on-going improvements in sensor technology. Considering this, but also the fact that our network is meant in itself as a research platform, the system design needs to be flexible to future upgrades. This is, in part, facilitated by the modular sensor node design we have opted for, as new system features can represent additional or interchangeable modules to the ones in the current system.

7. **Size reduction**

   A small size and a low weight are certainly desirable features for a sensor node. However, these objectives need to be traded-off against system flexibility and robustness.

8. **Minimizing cost**

   The cost of the platform has the ultimate impact on the number of nodes that the WSN will be comprised of for a given budget. Most of the design choices for the objectives above will have an impact on the end cost of the sensing platform. It plays an important role in component selection, particularly regarding sensor technologies.

4.2 Targeted Pollutants

In Chapter 1, we presented a list of the six common air pollutants included in the U.S. EPA criteria pollutants class. While covering as many pollutants as possible would be ideal, we need to take into account the cost design objective and select only the pollutants which are most relevant for the deployment site of our network: the city of Lausanne, Switzerland.

Therefore, considering that the levels of Pb and SO_2 have drastically fallen over the past decades, and that Lausanne is neither a heavy-industry city nor the site of any significant power station, we do not include them on the list of air parameters monitored by our sensor network. On the other hand, CO, NO_2, O_3 and PM are pollutants that are driven to a large extent by road traffic and are relevant for most urban environments. As such, we have included sensors for each of them in our design.

Additionally, we also target CO_2, which, while not on the list of pollutants directly relevant for health, is interesting in conjunction with CO, as the near-road CO/CO_2 ratio can be an indicator of the combustion quality of the vehicles in the city [52]. Furthermore, CO_2 is a non-reactive and non-toxic gas which can be safely used for running controlled experiments.

Summary

In this chapter, we presented and discussed the most significant design objectives for a successful mobile WSN for air quality monitoring. These include: measurement quality, coverage, robustness, energetic autonomy, maintainability, flexibility, size reduction, and minimizing cost. We then motivated our choice of targeted pollutants, which are CO, NO_2, O_3, PM, and CO_2.

5 Sensor Selection

ADVANCES in sensor technology over the past decade have lead to a significant increase in the range of miniaturized, relatively cheap sensors commercially available for measuring various air quality parameters. While this progress has fueled a new drive for non-traditional sensing applications, the measurement quality of these sensors is typically much lower than when using traditional instrumentation (e.g., absorption spectrophotometers, mass spectrometers, etc.), making the appropriate choice of sensors for a given application a critical task for the success of any monitoring system.

In this chapter, we provide a brief overview of the state of the art in sensor technology relevant to mobile air quality monitoring, while introducing also our own sensor selection. We start by discussing gas-phase chemical sensors, followed by particulate matter measurement systems.

5.1 Chemical Sensors

Commercially available gas-phase chemical sensors that are small enough to potentially be used in a mobile WSN application can be classified, depending on their transducer principle, as: semiconductor sensors, electrochemical sensors, Non-Dispersive Infrared absorption (NDIR) sensors, and Photo-Ionization Detector (PID) sensors. In this section we discuss these technologies and the sensors we selected.

5.1.1 Technologies

Each technology comes with its own set of advantages and disadvantages, and not all technologies can target all the pollutants we are interested in. Table 5.1 summarizes these technology specific characteristics. It is based on the work of Aleixandre et al. [53], and Snyder et al. [54], but also on our own experience evaluating various sensors.

Semiconductor sensors, also called metal-oxide or metal-oxide semiconductor sensors, are

Table 5.1 – Comparison of different chemical sensor technologies

Technology	Gases	Strengths	Weaknesses
Semiconductor	CO, NO_x, O_3, VOCs	Lowest cost; smallest size	Temperature, pressure and RH sensitivity; cross-sensitivity; fragile construction; high power consumption; frequent recalibration; slow response
Electromchemical	CO, NO_x, O_3	High sensitivity; very low-power; robust construction	No digital output; lower lifetime (1-2 years); cross-sensitivity; some sensitivity to temperature and RH; slow response
NDIR	CO_2	Stable; CO_2-selective	High power consumption
PID	VOCs	Very fast response	Most expensive; high power consumption; non-selective; frequent recalibration

the least expensive and smallest commercial chemical sensors currently available. The most used metal-oxide is tin dioxide (SnO_2), because of its high reactivity to a large number of gases [53].

The principle of operation is the following: When an oxidizing gas is present, the molecules of the gas react with the tin oxide trapping surface electrons and lead to an increase in the resistance of the SnO_2 layer. This reaction is reversible. For example, in the case of an O_3 sensor [55] this can be modeled as:

$$O_3 + e^- \rightarrow O^- + O_2 \tag{5.1}$$

where e^- is a conduction electron in the SnO_2 layer, and O^- is a surface oxygen ion.

Semiconductor sensors need a powered heater to increase the reaction rate of the sensitive layer. They have dynamic response times in the order of minutes, poor selectivity, and are affected by temperature and humidity variations.

Electrochemical sensors are composed of electrodes (two, three, or sometimes four) in contact with an electrolyte (mineral or organic acid). The two main electrodes, present in all electrochemical sensors, are called the working electrode and the counter electrode. Depending on the targeted pollutant, an oxidation or reduction of the targeted gas takes place at the working electrode. An opposite reaction (a reduction or an oxidation, respectively) will take place at the counter electrode. For example, for a CO sensor, the reaction at the working

electrode is:

$$CO + H_2O \rightarrow CO_2 + 2H^+ + 2e^- \tag{5.2}$$

and at the counter electrode:

$$\frac{1}{2}O_2 + 2H^+ + 2e^- \rightarrow H_2O \tag{5.3}$$

By composing Equations (5.2) and (5.3), we obtain the overall reaction of the electrochemical sensor as:

$$2CO + O_2 \rightarrow 2CO_2 \tag{5.4}$$

From Equation (5.4) it can be seen that the ingredients for the reactions are gases that are externally supplied to the sensor, and the product is an emitted gas, making the electrochemical sensor a catalyst, with none of its parts being directly consumed.

The electrochemical reaction at the working electrode (Equation (5.2)) produces a small current which can be measured by using an additional signal conditioning circuit. This circuitry is usually not supplied by the manufacturer, and, although example schematics are provided, it does imply a non-negligible integration effort.

Electrochemical sensors have very low power consumption and new generation sensors can be sensitive down to the ppm level. Moreover, their sensitivity can be improved through the use of chemical filters or the choice of electrolyte [53]. They are relatively more stable, requiring less frequent calibration than semiconductor sensors.

Non-dispersive infrared sensors are simple spectroscopic devices that use an infrared (IR) light source (a lamp) to illuminate the sampled gas. The gas in the sample chamber causes absorption of specific wavelengths. A detector then measures the attenuation on these particular wavelengths.

NDIR sensors are available for hydrocarbon gases and CO_2. They are non-selective for hydrocarbons, which share similar absorption characteristics. However, they are very selective for CO_2. The use of a lamp makes it relatively power hungry, and not well suited for low-power portable monitoring.

Photo-ionization detection sensors use high intensity ultraviolet (UV) light to ionize the molecules of chemicals inside a gas sample. The resulting ions are collected on electrodes within the detector cell, generating a measurable current.

PID sensors are very fast, having sub-second response times [56]. They are, however, the most expensive of the considered sensor technologies, have a high power consumption, and are not selective. Furthermore, the UV lamp needs to be regularly maintained and manufacturers recommend frequent re-calibration [53].
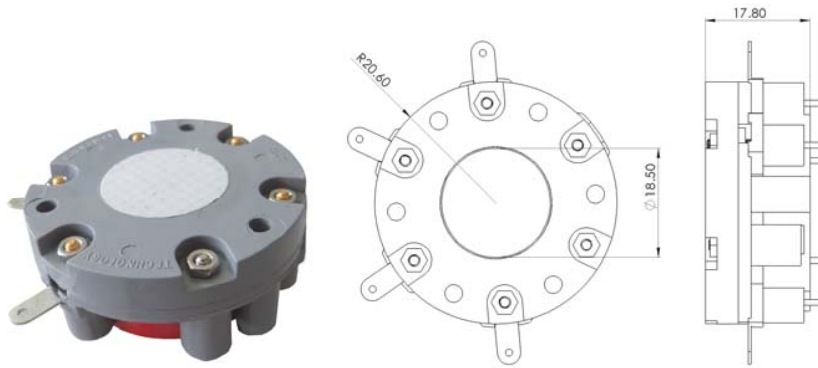
Figure 5.1 – City Technology A3CO Envirocel CO electrochemical sensor. (Note: all dimensions are in millimeters.)
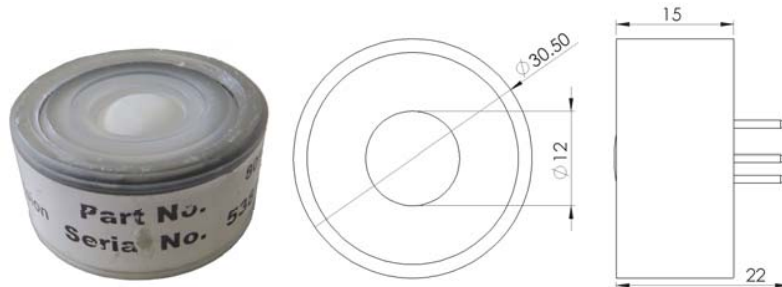

Figure 5.2 – KWJ Engineering NO2-SNL NO_2 electrochemical sensor. (Note: all dimensions are in millimeters.)

5.1.2 Selected chemical sensors

The choice of a particular sensor for an application is a non-trivial process that needs to take into account multiple criteria, including:

- **Performance** - Important performance characteristics for a sensor are sensitivity, accuracy, precision and response time. Ideally, judgment on performance should not be based only on producer specification but also on experimental validation.
- **Ease of integration** - Is the sensor integrated with all needed signal conditioning? Does it have a digital output? Is mechanical integration straight-forward?
- **Maintainability** - How frequent does re-calibration need to be done? How robust is it to varying weather conditions?
- **Cost** - The price of sensors can vary greatly depending on the targeted performance and additional features (e.g., signal preconditioning, digital output signal).

The final choice is usually a far from ideal compromise among these criteria. Sensor sensitivity is arguably the most critical, since a sensor is useless if it cannot detect the targeted pollutant at levels relevant to the application. Indeed, for our deployment, it outweighed by far other sensor characteristics, including fast response time.
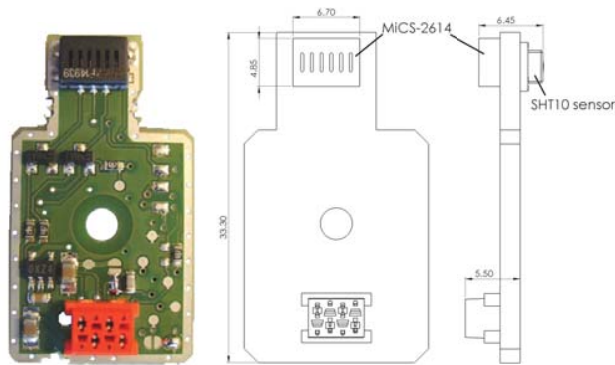
Figure 5.3 – SGX Sensortech MiCS-OZ-47 O_3 semiconductor sensor module. (Note: all dimensions are in millimeters.)
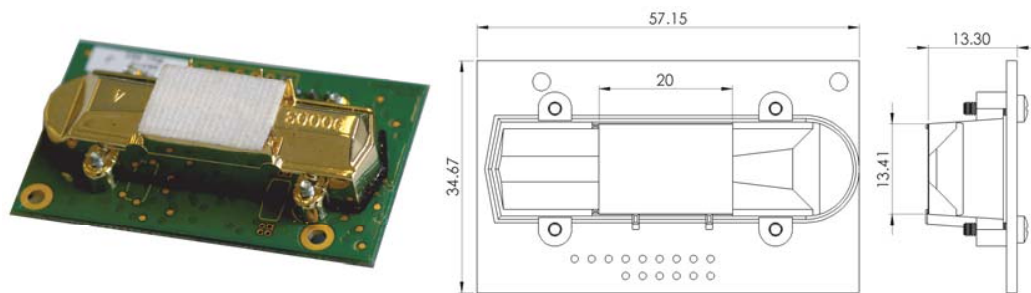

Figure 5.4 – General Electric Telaire 6613 CO_2 NDIR sensor module. (Note: all dimensions are in millimeters.)

The sensors we selected are the following:

- The **A3CO Envirocel** [57] from City Technology (see Figure 5.1) is a four-electrode electrochemical sensor for measuring CO. Its main features are a sensitivity of $0.2\,\mu A/ppm$, a nominal range of 0-500 ppm, high selectivity, temperature and humidity compensation (through the fourth, auxiliary electrode) and a response time of less than 40 s.

- The **NO2-SNL** [58] from KWJ Engineering (see Figure 5.2) is a three-electrode electrochemical sensor for measuring NO_2. Its main features are a sensitivity of -0.2 to $-0.5\,\mu A/ppm$, a nominal range of 0-10 ppm, and a response time of less than 60 s. It suffers from cross-sensitivity with O_3, SO_2, hydrogen sulfide (H_2S), and chlorine (Cl_2).

- The **MiCS-OZ-47** [55] from SGX Sensortech (see Figure 5.3) is a sensor module that integrates the MiCS-2614 semiconductor O_3 sensor [59], the Sensirion SHT10 temperature and humidity sensor [60], and a small 8-bit micro-controller. It provides a digital output over a RS232-TTL serial interface. It has a nominal range of 20-500 ppb, and a specified response time of 6 min. It is affected non-linearly by temperature and humidity.

- The **Telaire 6613** [61] from General Electric (see Figure 5.4) is a NDIR sensor module for CO_2 monitoring. Its main features are: analog and digital output (RS232-TTL and I2C), a nominal range of 0-2000 ppm, and a specified response time of less than 2 min. It also has an integrated self-calibration algorithm and a very long lifetime (up to 15 years).

5.2 Measuring Particulates

Traditionally, the most used metric for measuring PM is mass concentration (i.e. particle mass per volume of air). Other physical property metrics include particle number concentration and surface area concentration. Although desirable, not all of these metrics are usually available in a single device, and certainly not in one that would be small enough to lend itself to a mobile sensor network. When considering UFPs, particle mass concentration is usually an inappropriate metric because of their very low mass, which puts them below the detection limit of most measurement equipment.

For particle number and surface concentration, portable instruments capable of detecting from PM_{10} down to UFP-level are commercially available. These include *condensation particle counters* (e.g., TSI CPC models 8525 and 3007 [62]) and *electrical detection instruments* (e.g., Testo DiSCmini [63], Naneos Partector [64]). Although used in previous projects like AirCloud [39] or Data Canvas [65], the particle size range of small, low-cost optical dust sensors usually has a lower limit of $1\,\mu m$, making them insensitive to UFPs, which as mentioned in Chapter 1 is the most significant particle range to human health. For this reason we did not consider this technology in our selection process.

Condensation particle counters are generally more complicated to use and typically non-practical for integration into a long-term deployment compared to electrical detection instruments. They use a fluid (e.g., alcohol, water) to grow microscopic particles to a detectable size, which needs to be recharged frequently. Furthermore, device pose is important and tilting might cause measurement faults or even device failure. Finally, this device type only measures particle count, arguably a less relevant metric for health impact than surface concentration.

Electrical detection devices are more appropriate candidates for our targeted application. The Matter-Aerosol DiSCmini, described in detail by Fierz et al. [66], is a diffusion charging instrument that uses a high voltage pulsing corona to charge the incoming particles. The charged particles then flow through a diffusion stage, which captures the smaller particles, generating a diffusion current $I_{diffusion}$. The remaining particles are captured by a filter stage, generating a filter current I_{filter}. Based on the $I_{diffusion}/I_{filter}$ ratio, the average particle size can be computed, while the total current $I_{diffusion} + I_{filter}$ is related to the particle count. Furthermore, the diffusion current can be used to approximate particle surface concentration, or more precisely Lung-Deposited Surface Area (LDSA), a very relevant metric for health. The DiSCmini was used in a recent deployment in Zurich, Switzerland [67].

A more recent instrument, the Naneos Partector (see Figure 5.5), uses an even simpler measurement principle, by removing both the diffusion and filter stages, and replacing them with a Faraday cage [68]. The charged particles flowing through this cage induce a current which can be measured. This signal is then calibrated directly to the alveolar LDSA, providing the final instrument output. The simpler operation principle of the Partector compared to the DiSCmini leads to a smaller instrument size, but also to a much lower maintenance effort. In addition, the price of the Partector is significantly lower than the DiSCmini, which is an
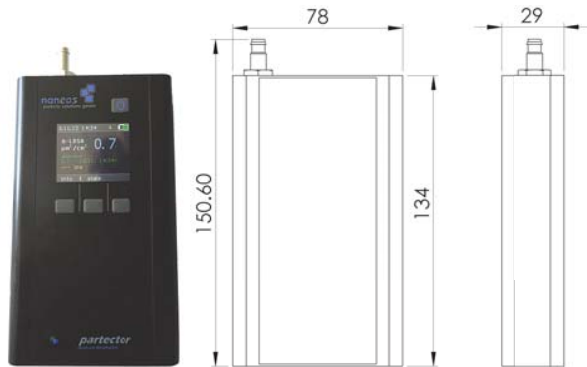
Figure 5.5 – Naneos Partector aerosol dosimeter. (Note: all dimensions are in millimeters.)

important decision factor considering that current PM instruments, that provide a sensitivity down to UFP level, are at least one order of magnitude more expensive than most chemical sensors. For all of these reasons, the Partector was selected for providing PM measurements in our deployment.

Summary

In this chapter we gave an overview of the sensor technologies which are relevant for mobile air quality monitoring applications. For gas-phase pollutants these are: semiconductor, electrochemical, non-dispersive infrared absorption, and photo-ionization detector sensors. For measuring PM down to UFP levels, electrical detection devices are the only viable option for long-term deployments. Finally, we also presented our own selection and the rationale behind our sensor choices.

6 Sensor Node Design

W E have developed a custom node design specifically for our project's scenario. This decision was prompted by both specific research questions and constraints of the anchoring vehicles. One of our goals was to develop a system that would be flexible to future mechatronic upgrades (e.g., additional sensor modalities, new sampling systems), and also to new mobility sources (e.g., private vehicles, cyclists). For this reason, a modular microcontroller-based system architecture was adopted, centered around a common communication bus.

6.1 System Overview

The computational workhorse for our sensor node is a 16-bit dsPIC33F micro-controller [69], a flexible device including digital signal processing and motor control capabilities. It is integrated in the DISAL R&D Board (see Figure 6.1), a fast development board that exposes most of the micro-controller's functions, which was developed within the Distributed Intelligent Systems and Algorithms Laboratory by Emmanuel Droz.

Each of the sensor node's logical modules uses one R&D Board together with a plug-in board that defines its functionality. The three plug-in boards that we developed for our sensor node are: a communication board, a localization board, and an air sampler board.

Figure 6.1 – The DISAL R&D Board - a flexible fast prototyping and development electronic board built around a dsPIC33F micro-controller.
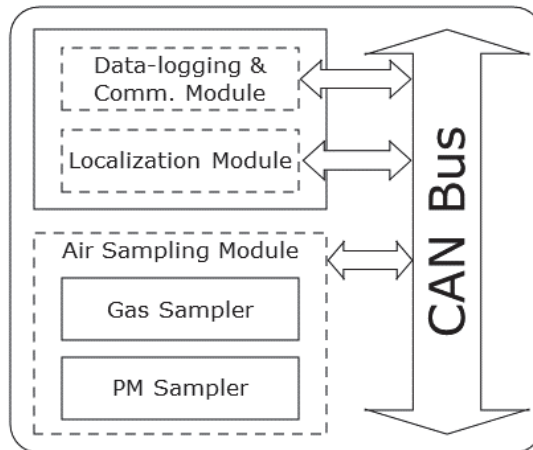
Figure 6.2 – Modular node architecture: dashed-outline boxes represent logical modules and solid-outline boxes represent mechanical modules.

The automotive standard controller area network (CAN) bus was a natural choice considering the noisy deployment environment. Using a simple twisted-pair implementation, CAN is a highly reliable multi-master bus which does not require specific device addresses. Instead, only messages carry a unique ID. This allows for a large number of devices to be connectable on a single CAN bus. While the resulting system is flexible to different configurations, the default configuration of our sensor node includes the following three components: a data-logging and communication module, a localization module, and an air sampling module.

A second layer of system modularity, regarding mechanical integration, was motivated by the need to minimize in-field maintenance time, and also by the vehicle specific constraints. While other deployments have used streetcars as anchoring vehicles [67], which have a rigid rooftop that can accommodate a monolithic node design, the roof of the Lausanne buses is made of flexible fiber. This reduced our options for anchoring points, but also meant that we needed to minimize the footprint of what we would deploy outside the vehicle. In early prototype phases, we were able to have all the equipment installed on top of the vehicle. However, as the complexity of the node grew, and especially after deciding to include PM measurements, we decided to keep on the roof exclusively what actually needed to be in open space (i.e. the air sampling module, and antennas). The resulting overall system architecture is schematically represented in Figure 6.2, and illustrates the dual nature of the system modularity.

6.2 Data-logging and Communication

The data-logging and communication module performs the functions of temporarily storing measurement data on the on-board SD card, and of managing the GPRS communication link to the back-end server. It is placed inside the same mechanical enclosure as the localization module (see Figure 6.3), which is stored in one of the vehicle's internal overhead compartments.
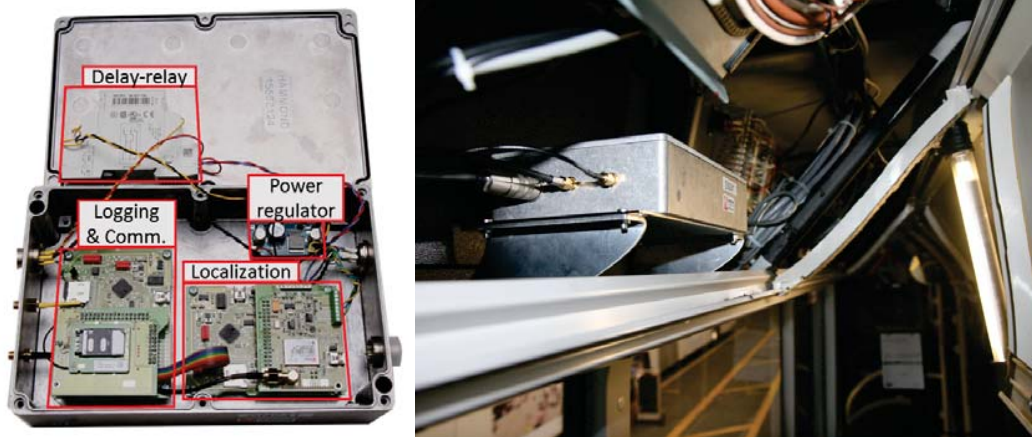
Figure 6.3 – Open enclosure containing the data-logging and communication module and the localization module (left). The communication and localization plug-in boards can be seen attached on top of their respective R&D Boards. The mechanical enclosure is anchored in one of the overhead compartments inside the bus (right).
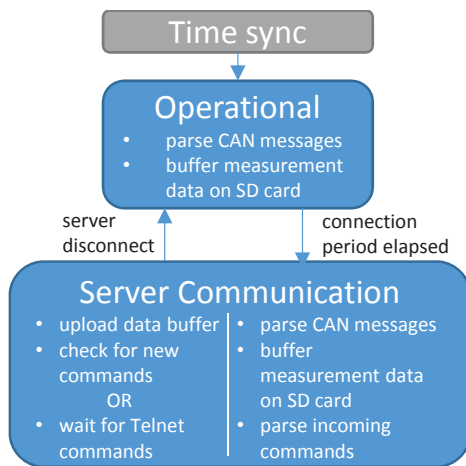

Figure 6.4 – Simplified finite state machine of data-logging and communication module operation.

This module acts as a finite state machine, which, after an intialization phase, alternates periodically between an *operational state* and a *server communication state*. A simplified diagram illustrating its operation can be seen in Figure 6.4.

While in the operational state, the module simply listens to the CAN bus and buffers the data received from the other modules on the SD card. When the connection timer fires (by default every 5 minutes), the module initiates the communication with the server. There are two types of server connections implemented. The default one is an automated communication sequence which starts with the sensor node uploading its data buffer. Once the data are successfully uploaded, the node queries the server for new commands for configuring the node operation (e.g., modify sampling rate, server connection period, server address, etc.). After executing all the server commands, the node disconnects and returns to the operational mode.
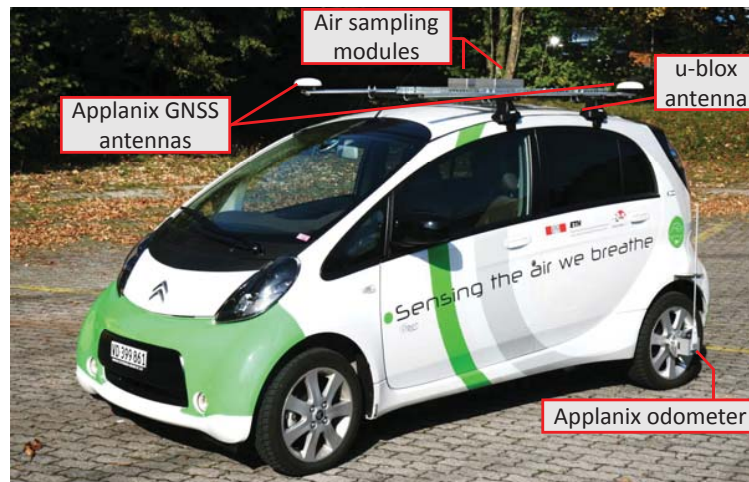
Figure 6.5 – Electric vehicle used as a flexible experimental platform.

The second type of connection is used for interactive, human-operated communication, through Telnet. In order for a Telnet communication session to be established, a server command for switching to this type of connection needs to have been sent to the node during the previous automated communication session. Telnet communication enables the operator to access the sensor node using a command line interface. It is mainly used for manual diagnostic purposes.

During both types of communication sessions, the logger continues to buffer incoming measurement data.

An additional feature of the module, particularly useful for system debugging and diagnosis, is the capacity to generate SMS alerts triggered by important violations of operational assumptions (e.g., watchdog timeouts, abnormal system reboot, brown-out, etc.).

6.3   Localization

The geo-referencing of measurement data is an essential functionality for a mobile sensor network. The use of Global Navigation Satellite System (GNSS) technology such as the Global Positioning System (GPS) or the GLObalnaya NAvigazionnaya Sputnikovaya Sistema (GLONASS), is an obvious choice for outdoor deployments. However, in an urban environment, multi-path effects due to the complexity of the landscape can significantly reduce the positioning accuracy obtained with low-cost receivers, leading to errors in the order of tens of meters.

GNSS positioning errors can be reduced through various methods, including map-matching algorithms, and sensor fusion with vehicle Dead-Reckoning (DR). Due to the relatively higher position uncertainty of buses compared to other public transportation vehicles (e.g., trams),

but also to the desire to keep the platform flexible for innovative position-driven sampling systems, we selected the automotive grade u-blox LEA-6R positioning module [70].

The LEA-6R module uses an Extended Kalman Filter (EKF) to fuse data from the GPS and DR measurements provided externally to the module. The options for DR inputs are vehicle speed-pulses, and yaw-gyro measurements. Also, an input for a digital temperature sensor is available, which can be used for temperature compensation of the gyro data. It includes an automatic on-line calibration algorithm for the DR input parameters (i.e. the speed pulse and the gyro scale factors, and the gyro bias).

In order to evaluate the performance of the LEA-6R performance, we performed a benchmark experiment, using it in parallel with the Applanix POS LV 120 system [71]. The Applanix uses two GNSS antennas, a high precision wheel-mounted Distance Measurement Indicator (DMI), an Inertial Measurement Unit (IMU), and Real Time Kinematic (RTK) corrections to achieve centimeter level accuracy. Both systems were mounted on the same vehicle (see Figure 6.5), and their data was logged as we drove on an approximately one hour track in and around Lausanne, which featured landscapes with varying degrees of complexity for the GNSS system (i.e. rural, sub-urban and urban). We conducted two such experiments, one using the DR information and the other one using only GPS data. The results of these experiments, as well as the localization tracks can be seen in Figure 6.6. In calculating the localization error, the Applanix system positioning was considered as the ground-truth.

The error distribution for these type of experiments will, of course, depend on the amount of time spent in each class of environment (i.e. regions with good sky visibility vs. regions with significant build-up). While not being an exhaustive module characterization campaign, the results of our experiments allowed us to conclude that using the DR feature on the u-blox module has a significant impact on limiting the localization error in areas without a clear view of the sky (e.g., downtown), which can get very large when using only GPS.

The position fix of the LEA-6R module is sampled at the maximum available frequency of 1 Hz and made available over CAN for logging. It includes latitude, longitude, speed, altitude, Horizontal Dilution of Precision (HDOP), and number of satellites used for the fix.

Apart from the LEA-6R, the localization module also includes a 3-axis accelerometer and a vehicle signal interface (see Figure 6.7). The latter provides the speed-pulse information needed for the LEA-6R, and also a rich set of vehicle context data useful for deriving mobility models and refining positioning accuracy. These data consist of information on the doors state, bus line, and current, next and final bus stops (see Figure 6.8). Furthermore, the vehicle signal interface provides the ignition key signal, which, together with a timed delay-relay (Figure 6.3), allows for a safe shutdown of the system when the vehicle is turned off.

Beyond the core role of providing positioning information, the localization module also supplies the real-time clock to the system, and exposes a precise 1 Hz time-pulse belonging to the LEA-6R, used for module synchronization.
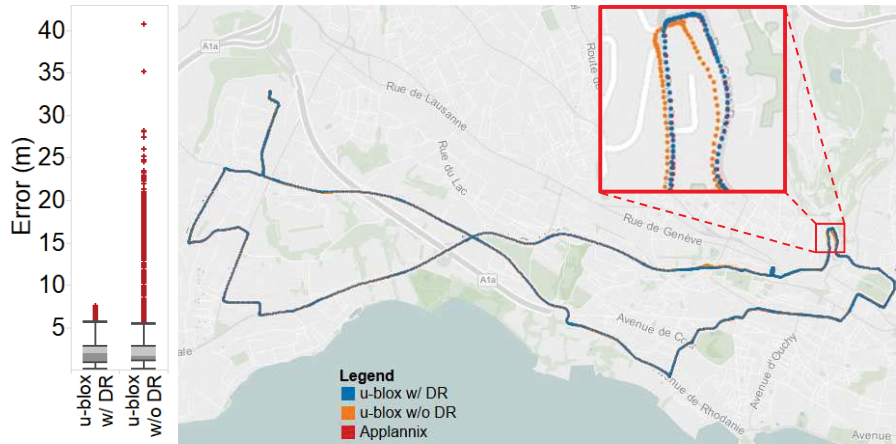
Figure 6.6 – Performance of LEA-6R module with and without DR: error distribution (left) and mapped experimental data (right). The detail on one of the regions in downtown Lausanne shows how the GPS-only positioning diverges from both the one using DR and the Applanix solution.
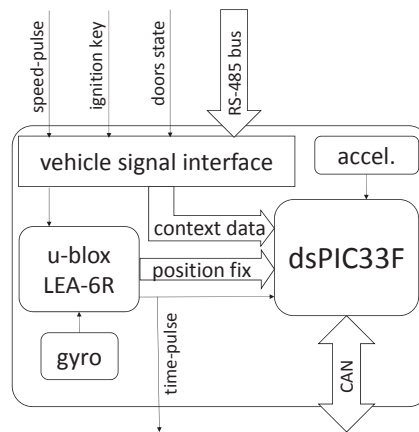
Figure 6.7 – Localization module data-flow: vehicle and positioning information is packaged by the dsPIC33F and broadcast over the CAN bus.

Figure 6.8 – Example set of vehicle context data.

Figure 6.9 – Air sampling module on roof of Lausanne bus.

6.4 Air Sampling

The air sampling module is the only part of our sensor node which was designed to be deployed outside the vehicle (see Figure 6.9). Since it is exposed directly to the varying weather conditions, it is also the most prone to failures. This raises the valid argument of whether having all the sensor node equipment inside the vehicle would be a better solution.

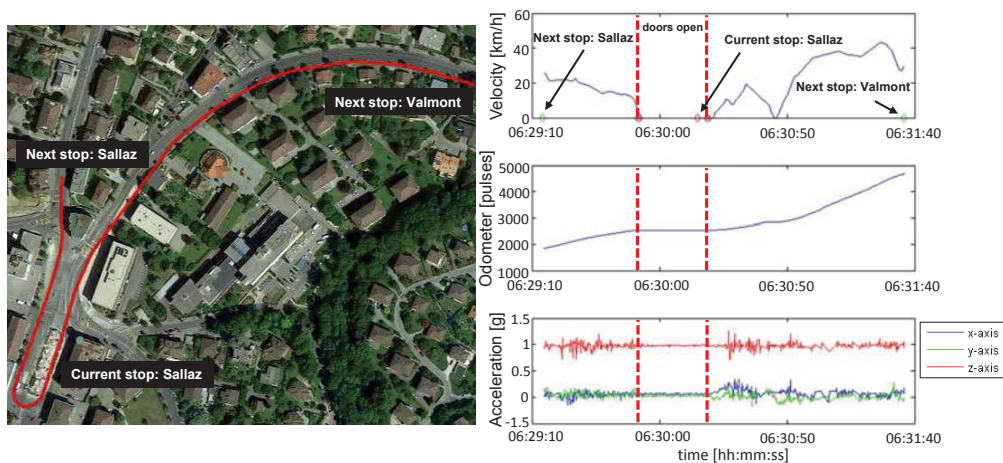Doing so, would imply the use of long tubing and means to generate air flow (e.g., a pump or a fan). Considering the fact that chemical sensors are generally very slow to react, using a long air transporting circuit would exacerbate this issue and would certainly need to be properly taken into account for a correct measurement-location association. Moreover, if this approach is taken, the task of choosing the right type of materials for the air circuit should not be neglected (e.g., employing polytetrafluoroethylene tubing, reducing metal surfaces), since reactive materials could render sensor measurements completely useless. One of the research goals of our deployment is to investigate the different axes of sampling system design, including active sniffing and closed sampling systems. In order to achieve this goal, we needed a basic design which would be flexible to future sampling system evolution. Designing the system from the start with the chemical sensors under a controlled air flow, inside the vehicle, would not have permitted the exploratory freedom that we needed. Finally, although having the sensor system completely inside the vehicle would reduce water exposure, it would not eliminate it completely, and provisions would still need to be taken to ensure that air inlets are splash and water protected.

For anchoring the air sampling module to the roof of the bus, we were asked by the public transportation operator to leverage the base of a now obsolete localization device that was present on all the vehicles we had selected. This 17×17 cm^2 welded steel platform provides us with an anchor to the rigid frame of the bus (Figure 6.10). Furthermore, it allows for all

Figure 6.10 – Exploded view of the air sampling module.

the cables to be securely passed through an opening in the center of the platform, leading to the anchoring cylinder piece. A switchbox was mounted on top of this opening to serve the double role of sealing the interior of the bus from water, and of providing a practical way of connecting cables. Since the load on the anchoring platform was going to be significantly higher than the obsolete localization system we were replacing, we also took care to reinforce its mechanical structure.

The support plate of the air sampling module is an additional platform that is fastened onto the vehicle anchor and enables an extension of the working surface, allowing the gas sampler and the particulate matter sampler boxes to be mounted side-by-side, as well as the GSM and GNSS antennas. Finally, a protective mask was designed to cover the two sampler boxes. Its role is to both provide additional mechanical protection, and to reduce temperature excursions due to direct solar radiation.

The gas sampler box houses the air sampling board, which manages all the air quality measurements of the module, including PM. It also contains all the selected chemical sensors (see Figure 6.11). In order to allow for future developments of the sampling system design, the sensors are mounted within a reserved 10×10 cm^2 planar surface, matching the corresponding opening in the support plate, and mounting points were foreseen for attaching additional air flow conditioning hardware. Extra volume is also reserved inside the enclosure for the same purpose. In its current state, the gas sampler is an open, passive measurement system that
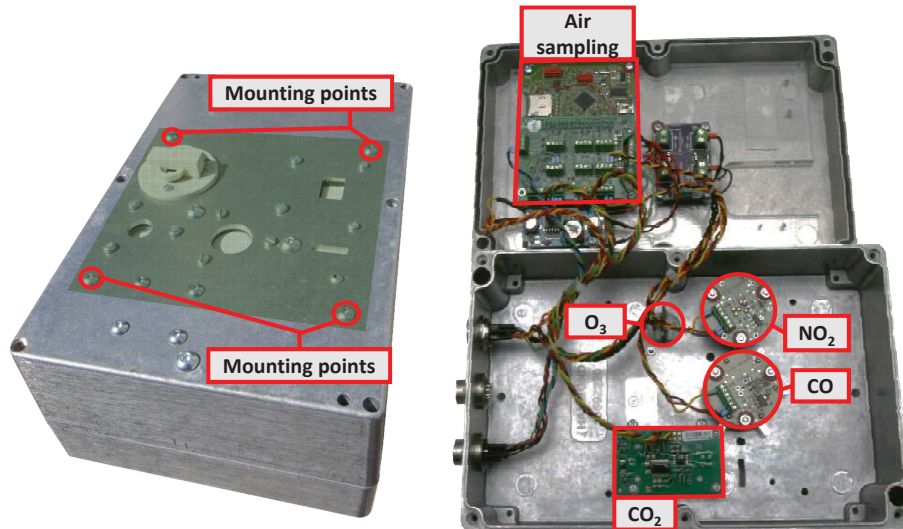
Figure 6.11 – Closed (left) and open (right) views of gas sampler box.

relies solely on movement induced flow, and passive diffusion when the bus is standing still. It is mounted with the surface exposing the sensors facing down, to protect them from direct rain.

All the chemical sensors, except for the O_3 one, have a planar surface which makes them straight-forward to integrate mechanically, using simple sealing joints (e.g., rubber O-rings). The O_3 sensor, on the other hand, has a non-traditional shape. The actual O_3 sensitive device is placed on one side of the sensor's PCB, while on the reverse side it has a temperature and humidity sensor. In order to attach it to the gas sampler box, we designed a protective two-part mask that would fit closely the shape of the O_3 sensor, while providing three planar mounting points to the surface of the box. The design was also intended to reduce water-splash exposure.

The CO_2 sensor, the O_3 sensor, and the Partector all have digital outputs and did not need additional electronics design for integration. However, the CO and NO_2 electrochemical sensors have only a small current output, which needs to be converted and amplified before being read by the microcontroller's ADC converter. The design of this signal conditioning circuit is typically a delicate task, since the large amplification needed to read the signal can lead to the final sensor reading having high levels of noise. Another important issue with electrochemical sensors is their rather long start-up transient. When initially powered on, their baseline value can take anywhere from minutes to hours to stabilize. In order to avoid this issue, and considering their very low power consumption, both the CO and NO_2 sensors are kept powered on all the time, using a different supply line than the one for the rest of the system.

The Naneos Partector is housed within a different enclosure from the chemical sensors. There are at least two reasons for this design choice. Firstly, the measurement principle of the
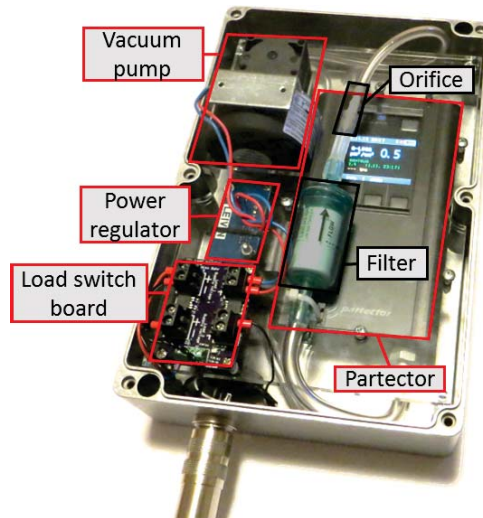
Figure 6.12 – View inside the PM sampler.

Partector is essentially different from that of the chemical sensors, with a controlled air-flow being a prerequisite for correct operation. Secondly, the specific assembly required for the integration of the Partector is relatively complex, which makes in-field debugging and repairs of the device unpractical for the most cases. In this context, having the PM sampler box separated permits more flexibility regarding gas sampler development, while reducing in-field maintenance, which is typically reduced to swapping the PM box.

The PM sampler assembly is shown in Figure 6.12. It includes a customized version of the standard Naneos device, which does not contain the default internal brushed motor pump, and also exposes a UART port for communicating with the air sampling module board. The reason for opting for an external brushless motor pump instead of the default one is that our long-term deployment is a very different scenario from that of portable personal exposure sensing for which the Partector was originally designed. Although typically smaller, brushed pumps need much more frequent maintenance compared with their brushless counter-parts. This makes them impractical for our deployment.

The basic air-flow circuit of the PM sampler starts with a conductive inlet tube leading to the Partector. The air is then pulled from the device by the external vacuum pump through a calibrated orifice, in a chocked flow regime, ensuring a stable flow around 0.5 lpm. In order to protect the pump, a filter is additionally fitted downstream from the Partector.

Although this can be modified on-line (e.g., through Telnet or server-based commands), by default, all the chemical sensors are sampled at 0.2 Hz, while the Partectors are sampled at their maximum rate of 1 Hz.

Summary

In this chapter we presented our sensor node design. It is a modular microcontroller-based system built around a CAN communication bus, with flexibility and maintainability as guiding principles. While the system allows for different configurations, the standard one includes: a data-logging and communication module, a localization module, and an air sampling module. For each of them, we presented their structure and functionality, and discussed our different design choices.

7 The Sensor Network

O UR mobile sensor network for air quality monitoring has been in continuous operation since autumn 2013. Beyond the issues of appropriate sensor selection, and sensor node design, the success of a WSN deployment relies also on network level design choices, and the capacity to maintain the system in operation, while understanding and improving on its initial limitations.

In this chapter, we introduce our mobile sensor network deployment, we discuss the server design, and the network maintenance. Finally, we also provide an insight into what to expect in terms of basic performance from this type of distributed mobile system.

7.1 Deployment

After a series of prototype single-vehicle tests starting in June 2011, the first final design sensor node was deployed on a bus on October 22, 2013. The rest of the vehicles were equipped at a steady pace, reaching the final number of ten buses by December 10 of the same year. The vehicle type used for all ten nodes is identical, consisting of the same model of articulated

Figure 7.1 – Equipped bus in operation in downtown Lausanne.

Figure 7.2 – Region covered by our mobile sensor network deployment.

Figure 7.3 – A sensor node equipped with two gas sampler boxes deployed in parallel at the Lausanne NABEL station.

diesel bus (Figure 7.1). Although it might seem tempting to use trolleybuses for this type of application since they have no emissions, trolleybus infrastructure usually covers only a subset of the public transport network, reducing significantly the achievable coverage. Regarding self-polluting the measurements, this is not a significant problem, considering the length of the bus and that the node is at the opposite end with respect to the vehicle exhaust.

The choice of the vehicle model was the result of discussions with the public transportation operator, who recommended it as the most flexible vehicle in use, being assigned to almost all the Lausanne lines. The resulting area covered is a large irregular region radiating from the urban center of Lausanne, through suburbs, reaching neighboring municipalities, and spanning an area of roughly 20×20 km^2 (see Figure 7.2).

Apart from the core of the deployment, represented by the bus nodes, an additional static node was deployed on the site of the local NABEL station, for calibration and sensor evaluation purposes (see Figure 7.3). Furthermore, another station was deployed on the electric vehicle introduced in Chapter 6 (see Figure 6.5), as a special purpose node of the network.

The first and most straight-forward purpose for this electric vehicle platform is its use as a
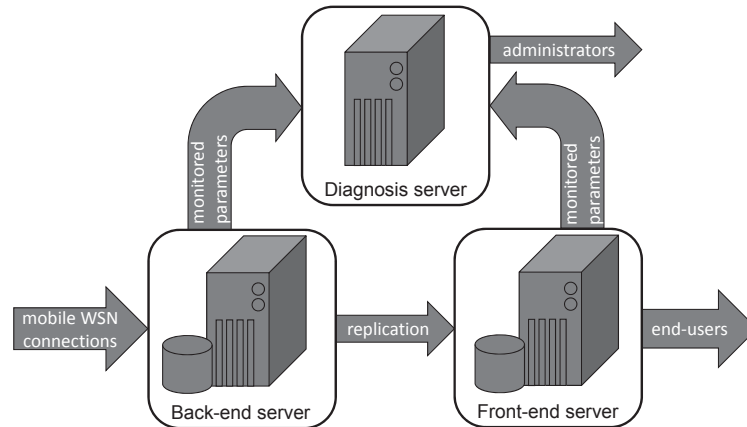
Figure 7.4 – Diagram of the server architecture.

test-bench for validating and improving the design of the sensor nodes. It provides a flexible platform for testing in parallel different design choices (e.g., to analyze the benefits and drawbacks of using an active sampling system).

Another purpose for this node is to perform targeted repeatable experiments to study topics such as the impact of mobility on chemical sensor measurements, or on-the-fly sensor calibration, but also to investigate regions of the city which are not reachable by the public transportation network.

7.2 Server Architecture

The measurement data stream generated by our mobile WSN is managed by a multi-server architecture, composed by: a back-end server, a front-end server, and a system diagnosis server (see Figure 7.4). Each of the three servers runs on a physically distinct machine.

All connections from the mobile WSN deployment are first treated by the back-end server, which parses the incoming messages, processes the data, and populates a PostgreSQL database with both raw and processed data. The processed data layers of the database get replicated on a front-end server, which can be accessed by the end-users. The end-users cannot access directly the back-end server. This ensures the protection of the critical process of data gathering and storage, which should not compete for system resources with user data requests.

On all three servers we have chosen to use the Global Sensor Networks (GSN) middleware [72]. GSN is a flexible, scalable, and platform-independent infrastructure for integrating heterogeneous sensor network technologies, which has been used in several real-world deployments [73]–[75]. It was developed at EPFL, by the Distributed Information Systems Laboratory.

It uses as its key abstraction the *virtual sensor*, which can be any kind of data source, including
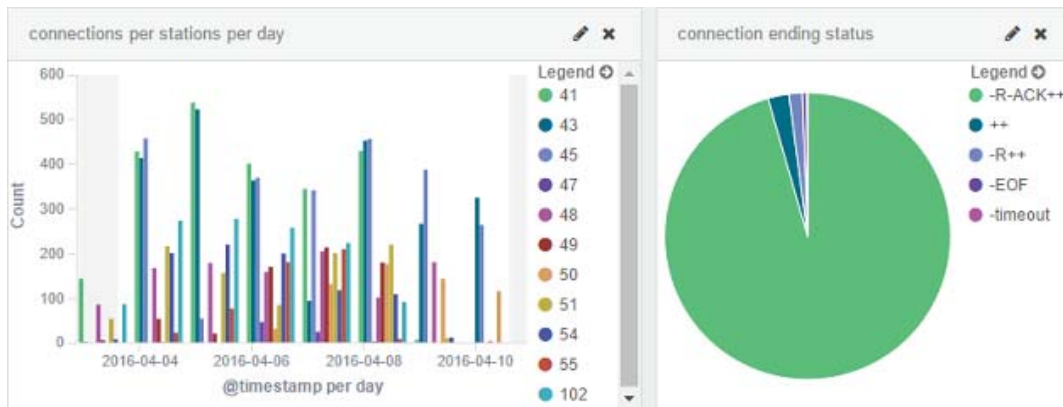
Figure 7.5 – The number of connections per sensor node per day over the past 7 days (left). Breakdown of communication disconnection types (right). Normal disconnects are depicted in green, while the other colors represent different types of erroneous disconnects. Note that stations 41, 43 and 45 were using a shorter connection period (every 2 min) at the time of this snapshot.

real sensors or combinations of other virtual sensors. Virtual sensors are hosted and deployed in GSN containers. Each container is independent and runs its own set of virtual sensors, although different containers or instances may interchange data in a peer-to-peer fashion.

Within a GSN instance, the virtual sensor manager administers a pool of virtual sensors, handling their life-cycles and managing the incoming data streams through wrapper interfaces. Once the data is captured by the wrapper, a processing layer can be programmed to store the observation data, annotate it, detect outliers, or apply corrections. Each of the three servers of our system runs a distinct GSN instance.

The back-end GSN instance receives the incoming sensor node connections and checks their integrity (e.g., checksum and time-stamp validation). Also, since the sensor measurements are not necessarily synchronized (e.g., the Partector measurements are periodic, but not triggered by the sensor node), the back-end GSN also geo-tags all measurements through time-based interpolation of the GPS stream. The front-end GSN instance replicates only the final, geo-tagged data.

For the diagnosis server, a newly developed GSN extension is used - a monitoring interface. Its purpose is to enable automatic diagnosis and facilitate system maintenance. Based on a predefined set of rules, the interface monitors the overall status of the virtual sensors (e.g., number of records produced, CPU cycles consumed, etc.), and calculates different anomaly metrics. The results of the metrics are then piped towards open-source software packages for easy visualization (e.g. Grafana, Kibana, etc.), and for generating alerts (using Seyren). Figure 7.5 shows an example of a visualization of system status data generated with Kibana.
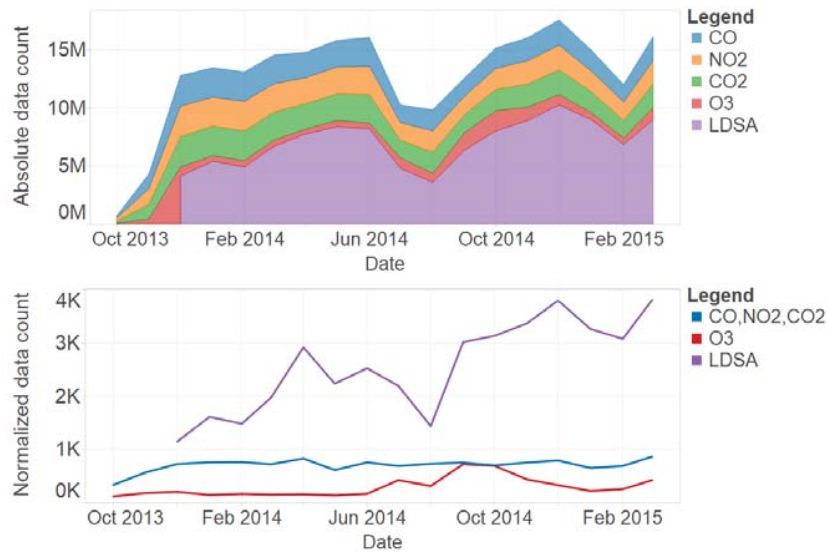
Figure 7.6 – Data throughput in number of measurement points over the whole network: absolute (top), and normalized by the number of hours in operation (bottom).

7.3 Maintaining and Improving the System

Efficient system maintenance plays an essential role in the success of a long-term outdoor WSN deployment. The key to achieving this lies in real-time monitoring of the health of the system. The most basic way of gauging the health of a deployment is by looking at how well it manages to perform the fundamental task of delivering measurement data. Sudden drops in data throughput can represent an indicator of system or subsystem failure. However, the maximum volume of data achievable with our deployment is modulated by the number of buses in operation at any given time, which varies as a function of vehicle maintenance needs, the dynamic bus-route allocation, and schedule variations (e.g., holidays vs. regular schedule). Figure 7.6 shows both the absolute data throughput and the throughput normalized by the number of hours in operation. Together they show a high-level view on the state of health of our deployment, with the first one being partly indicative of critical station failures, and the latter illustrating sub-system failures.

A **critical failure** can be defined as the type of system failure which, without an on-site intervention, breaks the data flow from a node irreversibly. These are typically rare events, but with significant consequences in terms of data loss if the detection and repair are not done rapidly. They are usually caused by SD card corruption, or more rarely SIM card or power supply failures.

Sub-system failures, on the other hand, can be more challenging to handle since, depending on the frequency of occurrence, they can be symptomatic of a problem of device integration. As Figure 7.6 shows, while the throughput of the CO, NO_2 and CO_2 sensors has been generally stable, the O_3 sensors had a significant problem of data loss. In fact, although the prior long
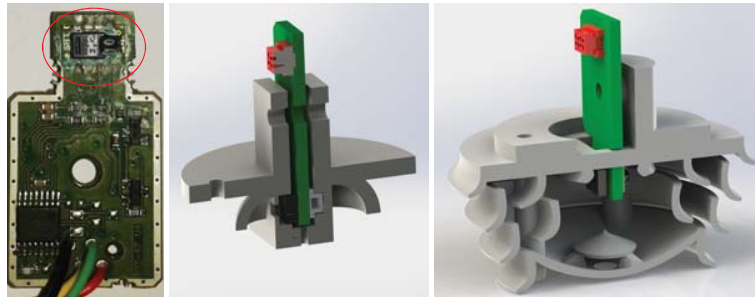
Figure 7.7 – Corroded O_3 sensor (left), old sensor mask (center), and improved design (right).

term tests of the sensor integration did not indicate any major issue, shortly after deployment they started to systematically fail, due to sensor corrosion (Figure 7.7).

The major difference from previous mobile tests of the O_3 sensor, performed from May until October 2013, was that the final deployment was done just before the cold season. For this reason, the main suspect was not precipitation, but condensation on the sensor, most likely happening when the buses returned to the significantly warmer depot. Moreover, the large vehicle access doors of the depot, which are usually open during the rest of the year, are kept closed to preserve heat during the cold season. This, however, leads to very high humidity because of poor ventilation of the vapors generated by the vehicle washing system, which is located within the same large hall as the indoor parking lot (where buses spend most of their offline time).

Our hypothesis was further re-enforced by the fact that the temperature and humidity sensor was getting corroded more than the heated O_3 sensor, being usually the first to fail. While being generally effective against water splash, the original design of the protective mask was probably too tight and any moisture due to condensation would be trapped inside with limited air-flow. To solve this issue, a new, more advanced mask design was produced in order to provide both additional protection against water splashes and air circulation around the sensor body. The new design was progressively installed for all the O_3 sensors starting from the winter of 2014-2015. No O_3 corrosion-caused failures have been reported since the upgrade.

Humidity is a major source of concern also for the PM measurement system. The Partector is sensitive to high humidity levels which can affect the measurement process, causing the data collected under these conditions to become meaningless. While not generally causing a direct failure, since the instrument returns to normal operation once it dries up, repeated exposure to high humidity levels can cause corrosion and weaken the high-voltage corona wire, which subsequently snaps. Moreover, since the PM system uses a pump to actively draw air into the Partector, in case of heavy water splashing this might lead to water being sucked into the device. Although the deterioration of the PM measurement systems was temporally less dramatic than for the ozone sensors, after over a year of operation, in the summer of 2015 all the PM measurement systems were taken down from the bus deployment, due to generalized Partector failures.

They have since been refurbished, and a number of new provisions were taken to improve their protection, including the use of heating resistors attached to the body of the Partector, a protective inlet similar in design to the new O_3 sensor mask, and a custom firmware upgrade done by Naneos for our deployment, with advanced health diagnosis. Since December 2015 we have started phasing the Partectors back into the deployment, with encouraging results so far.

Managing system failures can be an important man-hour budget sink, and steps need to be taken to streamline the maintenance process. To this end, a number of basic automatic diagnosis services have been implemented on the back-end server. These include automatic detection of critical failures (e.g., based on the time since the last connection or data transmission), and sensor health monitoring using simple statistics on the measurement distribution.

7.4   System Performance

While there are multiple criteria for evaluating the performance of a mobile network in monitoring a particular field, the most fundamental ones are its ability to capture the spatio-temporal variability of the targeted phenomenon, and the coverage it provides of the monitored area. In this section we present a small insight into the performance of our deployment from the perspective of these two basic criteria.

7.4.1   Spatio-temporal variability

In order to evaluate our network's capacity to capture spatial variability, we aggregate our measurements using a street-based discretization and proceed to a basic exploratory statistical analysis of the measurement data. We present our results for carbon monoxide and particulate matter measurements (i.e. LDSA).

Figure 7.8 shows the maps of the mean measured values for CO and LDSA, gathered from December 2013 to February 2015. They display an important spatial heterogeneity. As expected, we identify higher pollution levels in regions with more dense urban build-up and higher traffic. To quantify this relation, we perform a Pearson correlation analysis of the average measured CO and LDSA levels with respect to traffic volumes in different street links. We use traffic count data gathered at different locations around the Lausanne area during the measurement campaign of 2010 [76]. Figure 7.9 summarizes our results. For both CO and LDSA we obtain a statistically significant positive correlation between traffic volume and pollution levels, proving that our system is capable to capture at least to some extent the spatial variability of Lausanne.

In terms of temporal variability, our system is able to capture both weekly and daily pollution cycles. Figure 7.10 illustrates the daily and hourly average measured values aggregated for the entire measurement area, for the same time period.
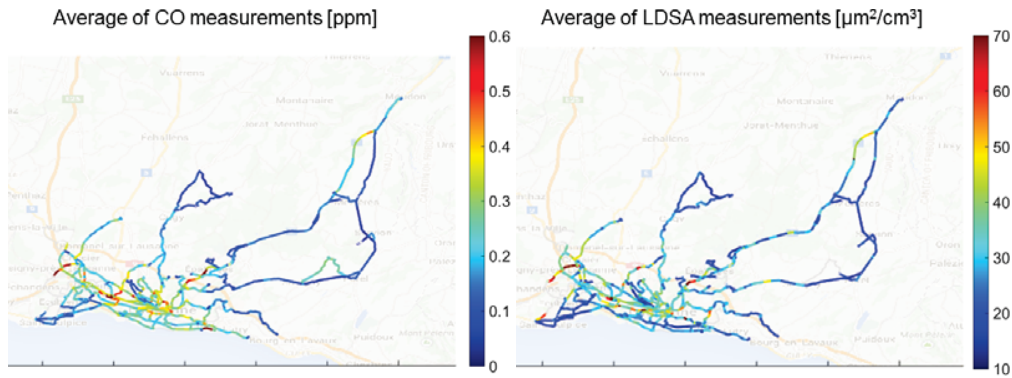
Figure 7.8 – Mean measured value maps show spatial heterogeneity.
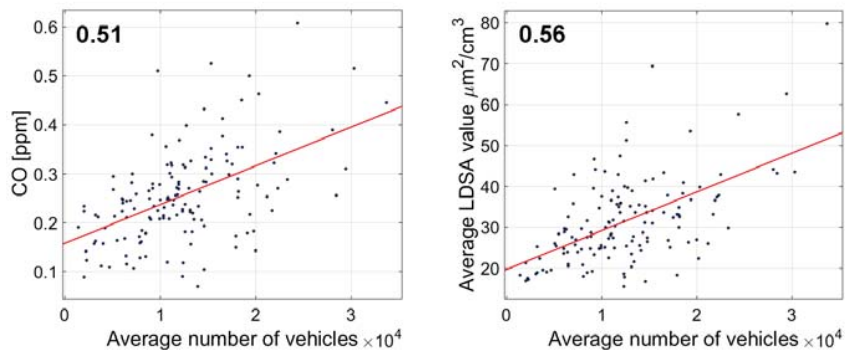
Figure 7.9 – Pearson correlation of mean CO (left), and mean LDSA (right) with traffic counter values.
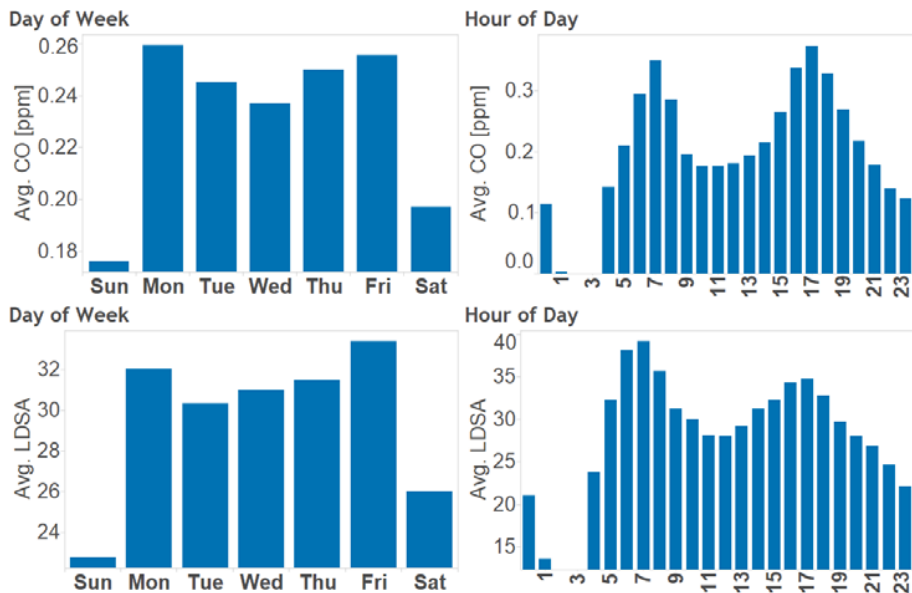
Figure 7.10 – Mean measurement values reflect lower levels on weekends (left), and rush hour peaks (right).
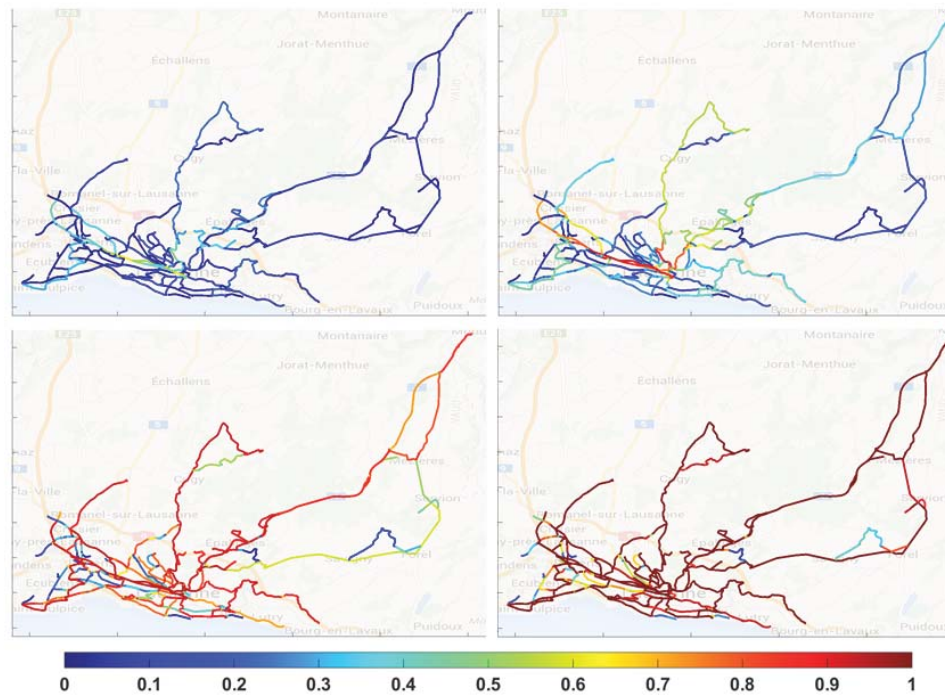
Figure 7.11 – Coverage probability of the network for hourly (top left), daily (top right), weekly (bottom left) and monthly (bottom right) time resolutions.
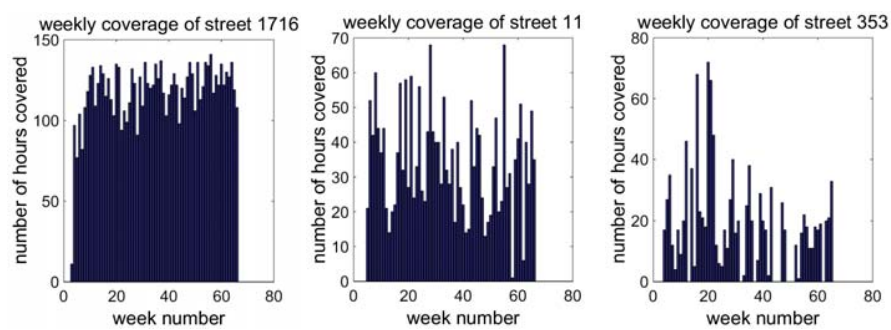
Figure 7.12 – The number of hours per week that one segment was covered. Segment 1716 is located near the bus-depot in Lausanne, segment 11 is downtown and segment 353 is in the suburbs.

7.4.2 Coverage

In our deployment, the uncontrolled movement of the nodes is unpredictable since the buses can be reassigned to different lines every few hours depending on the real-time needs of the public transportation operator. Therefore, the coverage of the network dynamically changes over time. Thanks to the large amount of measurement data gathered by our sensor network, we can nevertheless empirically calculate the network coverage.

We consider the following binary metric for defining coverage. For a given time resolution, a street link is covered if it is traversed at least once by a sensor node per time bin. With this definition, we compute the empirical probability of coverage for all links within our space domain, at hourly, daily, weekly and monthly resolutions. Figure 7.11 shows our results, based on data gathered within the first 18 months of deployment operation. As can be clearly seen, the probabilities are unevenly distributed in various segments, illustrating a dynamic coverage.

We have, furthermore, analyzed the time-wise distribution of coverage for the street segments. Figure 7.12 displays the number of hours per week that three exemplar streets have been covered by the network. The results show that for most of the down-town streets we have uniform coverage.

Summary

In this chapter we discussed the network level integration of our system. We first presented our server architecture, based upon the flexible and scalable GSN middleware, and the functionalities it provides. We then introduced our long-term deployment, in the Lausanne region, covering a $20{\times}20$ km^2 region. We discussed its maintenance, technical issues we encountered in the field, and our work in improving system reliability. Finally, we presented an overview of basic system performance, in terms of its ability to capture spatio-temporal variability, and to provide coverage of the monitored area.

8 Conclusion

O ver the past decade there has been a growing interest in using mobile wireless sensor networks for bridging the gap among traditional air quality measurement systems (i.e. static monitoring systems and portable devices). While recent advances in sensor technology have brought to market an increasing number of small and relatively cheap sensors, the process of integrating them into a successful monitoring deployment is fraught with significant scientific and technical challenges.

In our work, we have developed and deployed a mobile wireless sensor network for monitoring multiple air quality parameters in the city of Lausanne, Switzerland, by anchoring our sensing platforms on local public transportation vehicles. The deployment is in full operation since November 2013, and is planned to be maintained for a minimum of three years.

We shared the experience we gathered in designing, deploying, and exploiting this type of mobile sensor network. We discussed design objectives and described the development process step-by-step, from the selection of targeted pollutants, to the choice of sensors, sensor node and server design, deployment and maintenance. Furthermore, we provided an insight into the performance of the developed system, in terms of coverage and its ability to capture the spatio-temporal variability of the urban pollution landscape. We hope that it can constitute an integrated guide to researchers and developers interested in this topic.

8.1 Discussion

We developed our mobile sensor network with two main research goals. First, we were interested in investigating the opportunity of using mobile sensor networks for bridging the distance between traditional methods of air quality measurement. The second goal is particularly unique to our work, as we wanted a system that could represent a flexible research platform for exploring mobile air measurement system design. We consider this research direction essential to achieving reliable accurate mobile air quality monitoring.

To this end, rather than opting for a monolithic design, as the ones used in other related work [42], [67], we have developed a modular sensor node architecture targeting a diverse and expandable list of pollutants (e.g., CO, NO_2, CO_2, O_3, particulate matter).

Through a core of 10 bus-mounted sensor nodes, and some additional special purpose stations, our network has gathered hundreds of millions of measurement points. In order to manage this massive data, we have adopted GSN for our server architecture, a middleware solution designed to handle the life-cycle of virtual sensors (i.e. devices, objects or people observing properties around them). This includes sensor data acquisition, metadata management, storage, real-time processing and publishing of measurement data.

8.2   Outlook

Despite being a very powerful research platform in its current form, there are, of course, still a number of improvements that could be made to our mobile sensing system. On the sensor node side, we are interested in studying the benefit of adding wind sensing. The ability to gather local wind data would be a very interesting enhancement of the basic node, since wind conditions have a large impact on air pollutant dispersion.

Another desirable improvement would be to endow the sensor node with the possibility of wireless and remote firmware reprogramming. This might however imply a change of architecture at least at the level of the logging and communication module (i.e. the use of at least a light-weight operating system).

Finally, in order to simplify maintenance, which for a system of this size is in no way negligible, more work needs to be done in implementing automatic system health diagnosis, going beyond the currently basic fault-detection functionalities.

Summary

We conclude our work on designing a mobile air quality monitoring network by stating our goal for this written account of our development and long-term deployment experience to serve as a helpful integrated guide to researchers and developers interested in this topic. We discuss the specific research goals that guided our most important design choices and, finally, we present some possible directions for further improvement of the system.

Mitigating Mobility Effects on Chemical Sensor Measurements

Part III

9 Introduction

THE main advantage of using mobility stems from the possibility of extending spatial coverage for a given number of sensor nodes. Nonetheless, the inherent dynamic nature of mobile coverage represents a significant challenge for attaining spatio-temporally complete high-resolution air pollution maps. This issue can be addressed through statistical modeling techniques, as suggested by Hasenfratz et al [67], Li et al [77], or Marjovi et al [51]. However, all of these works have considered solely particulate matter (PM) measurements, and to the best of our knowledge, no similar results have been published for gaseous pollutants using mobile measurement data.

The reason for this is that, differently from electrical detection devices used for PM measurements, the current commercially available chemical sensors for measuring gas-phase pollutants suffer from a number of issues that need to be solved before the aforementioned modeling methods can be applied. These include temporal drift due to sensor aging, cross-sensitivities to other chemical components or environmental parameters (e.g., temperature, humidity, pressure, etc.) [53], [54], low SNR at ambient concentration levels, and slow response times, which, in the context of mobility, leads to significant measurement distortion.

The problems of sensor drift and cross-sensitivity can be addressed through on-line calibration methods, and, although still an open problem, a significant body of work has been building up on this topic in recent years [49], [78]–[81]. The issue of signal distortion due to the interplay of slow sensor response and high platform mobility has, however, received little attention.

9.1 Motivation

Technological advances in the field of chemical sensors have significantly increased the availability of small and cheap sensors for measuring various gas-phase species. Nevertheless, the data quality obtained from these sensors is significantly lower than that of traditional instrumentation (e.g., absorption spectrophotometers, mass spectrometers, etc.).

The most commonly used chemical sensors for WSNs applications use either electrochemical or semiconductor technology. Both classes of sensors suffer, although to different degrees, from temporal drift, cross-sensitivity (to other gases or environmental parameters), and long response times [53], [54]. The first two issues can be addressed through online calibration methods. This subject will be discussed in depth in Part IV.

Beyond calibration, dealing with the slow dynamics of chemical sensors in a mobile application is a key issue. Chemical sensors typically have long response times to stimuli, which can range from a few seconds to multiple minutes. While for static deployments, this issue can be largely neglected, for mobile platforms it can induce significant distortion of the measured signal with respect to the underlying concentration levels.

Within the field of mobile WSNs, the slow-dynamics of chemical sensors has, unfortunately, so far gone unaddressed: most of the previously mentioned works assume measurements to be instantaneous. This problem has, nevertheless, been acknowledged in the field of mobile robot olfaction. Strategies for mitigating such effect include speed limitation of the robotic platforms [82], cycling between movement and stationary measurement behaviors [83], and the use of specially designed air sampling systems [84]–[86].

While the first two strategies cannot generally be considered in the case of mobile WSNs, which typically leverage parasitic mobility, the design of sampling systems needs to be investigated.

9.2 Experimental Set-up

In order to isolate and investigate the effect of mobility on a chemical sensing platform, we designed a controlled experimental set-up. As pointed out also in the work of Trincavelli et al. [87], who proposed an experimental set-up for investigating chemical sensor dynamics for odor discrimination, this task is in no way trivial due to the stochastic nature of gas dispersion, and the difficulty of getting a ground-truth of the measured field.

We decided to conduct our experiments inside a wind tunnel, which allowed us to control both the wind and movement speeds, providing experimental repeatability. Thanks to upgrades to our wind-tunnel infrastructure, we were able to improve our experimental set-up during our studies, going through two similar, but distinct configurations. For the sake of clarity, we will henceforth refer to the initial experimental set-up as *configuration A*, and the newer version of the set-up as *configuration B* (see Figure 9.1).

For all experiments we used the gas sampler module we designed for our real-world deployment (see Figure 6.11). Out of the list of available sensors we opted to focus on the City Technology A3CO carbon monoxide electrochemical sensor. The reason for this choice is based on the high selectivity and relatively good temporal stability (i.e. low drift) of this sensor, which permit a better isolation of the problem targeted in this study.

In both experimental configurations, the sensing module was attached to a traversing system
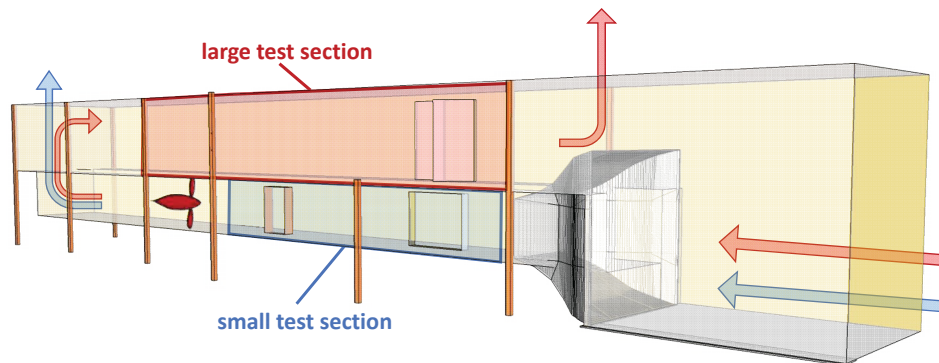
Figure 9.1 – Wind tunnel experimental configurations: configuration A used the small test section, with the air circulating according to the blue arrows in the figure. For configuration B, we used the larger test section, with the corresponding air circulation outlined by the red arrows.
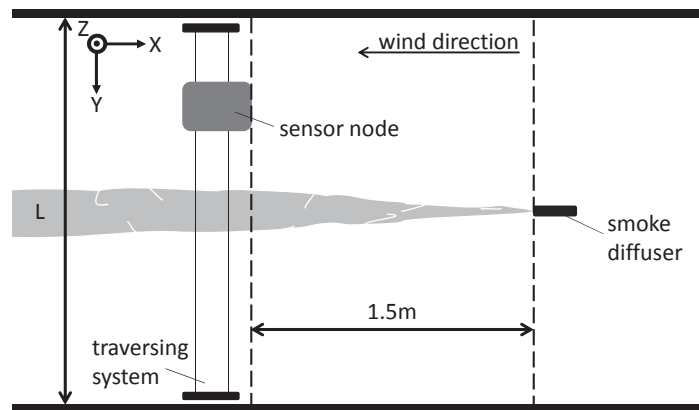
Figure 9.2 – Top view of experimental set-up inside wind tunnel (not to scale). For configuration A, the value of L is 2 m, while for configuration B, it is 4 m.

(i.e. a cartesian coordinate robot), which allowed us to vary the movement speed in a controlled and repeatable way. As a source of pollution, we used a smoke machine which created a chemical plume that our sensor node is able to detect. Figure 9.2 shows a schematic representation of our experimental set-up.

9.2.1 Wind tunnel

We performed our experiments in a boundary layer wind tunnel, with a maximum wind speed of 24 m/s, achievable in the smaller test section. For configuration A we used the smaller test section, which has a $10 \times 2 \times 1.5$ m^3 working area. The dimensions of the larger section, used in configuration B, are $19.5 \times 4 \times 1.95$ m^3. The features of the small test section are higher maximum wind speed and more laminar flow. However, the larger test section permits higher traversing system velocities because of its bigger cross-section, which permits accelerating up
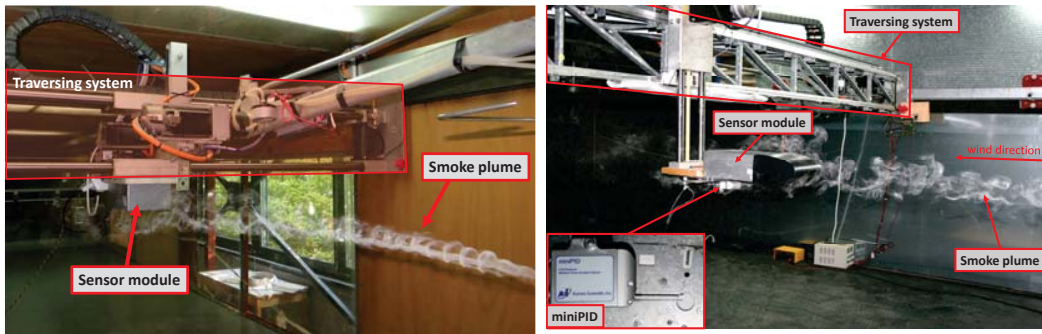
Figure 9.3 – Snapshots of our sensor module traversing the smoke plume in experimental configuration A (left) and configuration B (right). The location of the miniPID device used as ground truth for configuration B is outlined in the corresponding picture.

to 1.25 m/s (compared to approximately 0.5 m/s in the lower channel).

In both configurations, the sensor box was attached with the sensors facing down (similar to our bus deployment), onto the respective section's traversing system. During our experimental runs, we performed controlled movements in the cross-wind direction (on the Y-axis of the traversing systems), at different levels of the Z-axis, while keeping the X-axis locked.

9.2.2 Chemical plume source

For generating a chemical plume we used the Pea Soup Wind Tunnel Air Flow Tracer SGS-90 [88], which produces smoke by heating a white mineral oil flowing through a diffuser. The quantity of smoke can be controlled by tuning the oil flow and heating voltage, and was set to be consistent for each of our different experimental sets. The width of the plume is modulated by the background wind speed.

A great advantage when using a smoke machine as plume generator is that it permits a direct visualization of the shape and stability of the plume. Snapshots of our two experimental configurations with clearly visible smoke plumes are showed in Figure 9.3.

9.2.3 Ground truth

The method for estimating the plume ground truth is different for the two configurations. For configuration A, we designed and conducted the following experiment: we commanded the traversing system to move at regularly spaced positions on the Y-axis, with a 5 cm increment. At each of these positions 130 s of data were acquired (i.e. more than the manufacturer stated response time of maximum 40 s). We used a sampling period of 60 ms, which was kept constant also for the rest of the experimental work presented in this chapter. After each interval of static data acquisition, the sensor was moved out of the plume (at the zero position), and left for 120 s to ensure its full recovery, before moving to the next 5 cm increment. The data was then
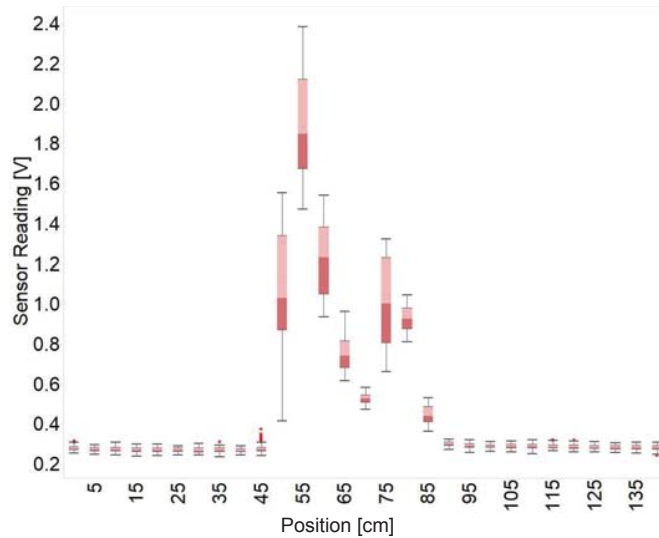
Figure 9.4 – Cross-wind plume profile ground-truth estimation. The boxplots of the data acquired at each location illustrate the local distribution of the gas concentration. The structure of the plume was kept stable throughout all experiments.

processed by clipping the first 40 s of data for each position.

The resulting estimation of the plume profile is represented through a collection of boxplots in Figure 9.4. In this manuscript we assume the convention of extending boxplot whiskers up to 1.5 of the interquartile range. The resulting graph shows that, in spite of an important variability of the concentration level within the plume, our experimental set-up is able to maintain its overall structure and position. The distinct two-peak concentration profile is most probably due to the fact that, at the considered wind speed, the smoke diffuser induces a ring-like cross-section to the plume. This effect can be directly observed in the shape of the corresponding smoke plume in Figure 9.3.

The most significant improvement on the experimental set-up brought by configuration B is the elimination of this rather cumbersome and very time-consuming method for acquiring the ground truth. This was done by the integration into our experimental set-up of the fast response 200B miniPID measurement system [56] from Aurora Scientific, a photo-ionization detector. This was attached in close vicinity of our electrochemical sensor (see Figure 9.3). Both the miniPID and the electrochemical sensor are sensitive to the plume generated by the smoke machine.

Summary

In this chapter we introduced the problem of measurement distortion in mobile sensor networks that use low-cost chemical sensors. This is due to the slow dynamic response that is characteristic for both electrochemical and semiconductor sensors, the most commonly used sensor technologies for mobile air quality monitoring. The long response times of these sensors in conjunction with the high mobility of anchoring platforms leads to significant signal distortion which has, unfortunately, so far been neglected in previous works. We propose an experimental approach to the study of this problem. For this purpose we designed a rigorous experimental framework using a wind tunnel facility.

10 Proposed Techniques

W<small>E</small> study the impact of sensor dynamics on measurement accuracy and locality through systematic experiments in the controlled environment of a wind tunnel. We then propose two methods for mitigating this problem: (i) estimating the underlying true signal using a sensor model and a deconvolution technique, and (ii) enhancing the measurement quality of chemical sensors through the use of active sampling.

The results presented in Sections 10.2.1 and 10.2 were performed using the experimental configuration A, while the results in Section 10.3 were obtained with the upgraded set-up (i.e. configuration B).

10.1   Mobility-caused Measurement Distortion

In order to illustrate the distorting effect of mobility, we conducted the following experiment: we performed a series of repeated dynamic scans of the plume at several constant speeds. Once the sensor node reached the end of a run, we continued logging the data until the sensor recovered its baseline reading. We considered movement speeds of 2.5 cm/s, 5 cm/s and 10 cm/s, and performed 10 runs for each setting. The background wind speed was kept constant at 0.5 m/s for all experiments.

The results of this set of experiments are presented in Figure 10.1. These show that, even for relatively low speeds of the sensor node, the signal distortion is significant. Due to its slow response, the chemical sensor appears to perform largely as an integrator of the underlying signal over time, leading to a larger distortion when the speed of the node is increased.

A simplifying analogy, which does not take into account the perturbing effect of the sensor box on the flow, can be made with motion-blurring in photography, which happens when the exposure time of a camera system is long relative to its movement speed. A typical signal processing approach for reducing the effect of motion-blurring is through deconvolution.
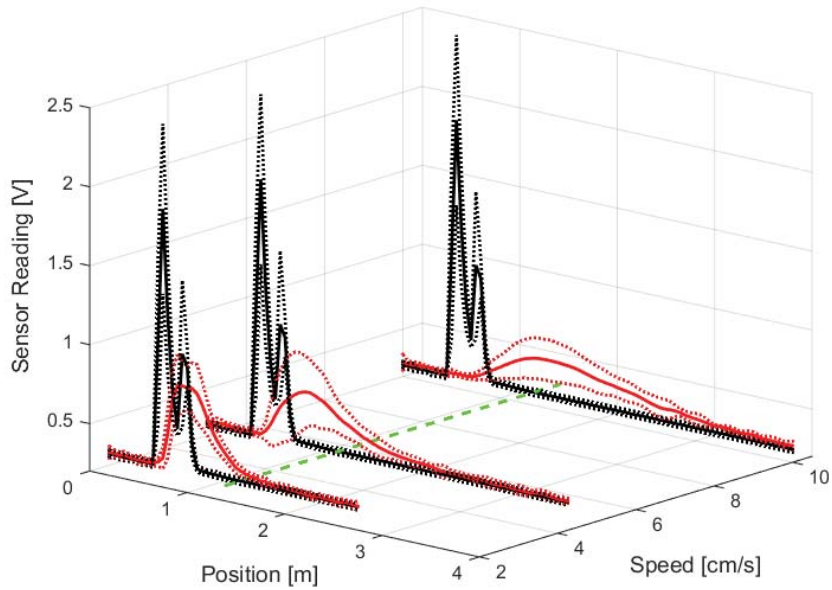
Figure 10.1 – Average measured signal (red) over 10 runs at different speeds of the traversing system versus the reference plume profile (black). Dotted lines represent 2σ confidence intervals. The green dashed line marks the end of each run where the sensor node stops moving. For visual clarity, we unfold the data recorded at this position as if the movement would have continued with the same velocity.
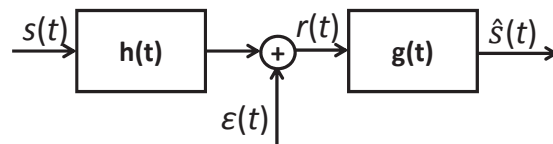
Figure 10.2 – A deconvolution procedure convolves a distorted signal with a filter $g(t)$, in order to obtain an estimation of the original signal.

10.2 Signal Reconstruction through Deconvolution

We adopt the following formalism: Let $s(t)$ be the underlying carbon monoxide concentration level that we would like to measure, and $h(t)$ the impulse response of the electrochemical sensor we are using, and $S(\omega)$ and $H(\omega)$, their respective Fourier transforms. The sensor reading $r(t)$ can then be expressed as:

$$r(t) = (s * h)(t) + \epsilon(t) \tag{10.1}$$

where the symbol $*$ denotes the convolution operation, and $\epsilon(t)$ represents additive noise.

The goal of a deconvolution algorithm is to find a filter $g(t)$ ($G(\omega)$ in frequency domain), which

applied to the measured signal, can achieve a good estimation of the original signal:

$$\hat{s}(t) = (g * r)(t) \tag{10.2}$$

A schematic representation of the principle of deconvolution is shown in Figure 13.7.

Depending on the choice of the independent variable for our signals, the measurement distortion and dencovolution filter can be studied either through space, or through time. These approaches are equivalent as long as the sensor node moves at a constant speed. However, since this is not a valid hypothesis for real-world mobility, we opt to conduct our analysis of the signals as functions of time. This allows us to have a single model of the sensor's distortion process, independent of changing velocities.

A popular method for performing deconvolution is through the use of the Wiener filter, which minimizes the mean squared error between the estimated and the desired signals. The Wiener deconvolution filter is most easily described by using the Fourier transform:

$$G(\omega) = \frac{H^*(\omega)}{|H(\omega)|^2 + 1/SNR(\omega)} \tag{10.3}$$

where the superscript $*$ symbol represents complex conjugation, $H(\omega)$ is the Fourier transform of $h(t)$, and $SNR(\omega)$ is the signal-to-noise ratio. In the ideal case, when noise is absent, the Wiener filter becomes the inverse of $H(\omega)$. We worked under the assumption of white noise (i.e. having a constant power spectrum).

10.2.1 The sensor model

In order to use the Wiener filter, a model of the chemical sensor is required, and also an estimation of the noise present in the measured signal. We assume a second-order damped linear time invariant LTI system model for the chemical sensor, of the form:

$$H(\omega) = \frac{A_0}{(1 + j\omega\tau_0)(1 + j\omega\tau_1)} \tag{10.4}$$

where A_0 is the zero-frequency gain of the model, and τ_0 and τ_1 are its two time constants. This model was selected, after multiple trials with other model orders, as it provided a good balance between complexity and data-fitting performance.

We estimate the model parameters through system identification, using the dedicated Matlab Toolbox. The reference average plume profile was used as system input data, and the set of measured signals over the set of dynamic experiments presented in section as system output.

The SNR was estimated by using the sensor measurements outside of the plume, where we assume that the variation of the sensor reading is caused only by additive noise. More specifically, we calculated it using the `snr` routine, included in the Matlab Signal Processing
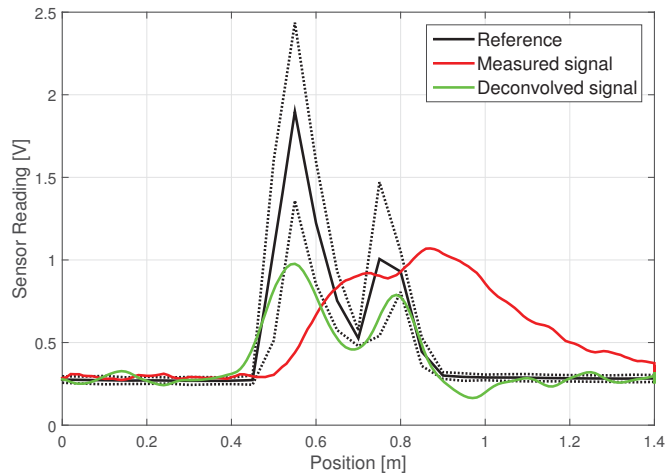
Figure 10.3 – Signal deconvolution example. The measurement data was gathered during a run at 2.5 cm/s.

Toolbox. This requires an estimate of both the noise and the clean signal. Based on the assumption that, the measurements outside of the plume are only due to noise, we applied a zero-phase smoothing filter, varying its span until the amplitude of the residual error outside of the plume was below a fixed threshold for all experiments. For all the results presented in this chapter this threshold was set to 10 mV. The smoothed result provided the estimate of the clean signal. The estimation of the noise was obtained by subtracting the estimated clean signal from the raw data series. The advantage of this method is that, once the appropriate span for the smoothing filter is identified inside the wind tunnel, it can also be used for estimating the SNR of outdoor data (where identifying the signal baseline would be very difficult). This proved useful in the case of the outdoor experiment presented in Section 10.3.4.

In order to identify the sensor model, the experimental data was divided into model training and evaluation sets, through a leave-one-out cross-validation scheme. Each training set contained measurement data for all of the considered speeds. After training all the models (i.e. 10 models per considered speed), we applied the Wiener deconvolution to each experimental run, using the model for which the respective data was not included in its training set. Figure 10.3 shows an example of the signal estimation through deconvolution for one of the experiments.

10.2.2   Results

For evaluating the performance of the deconvolution for mitigating the effect of the chemical sensor's slow dynamics, we have used the following two metrics:

- **Feature localization accuracy** - This metric quantifies the ability of the system to estimate the position of discrete features in the measured field (e.g., chemical leaks, pollution hot-spots). It implies a threshold segmentation of the measured and reference signals, to define what constitutes a feature (see Figure 10.4). For a pair of features
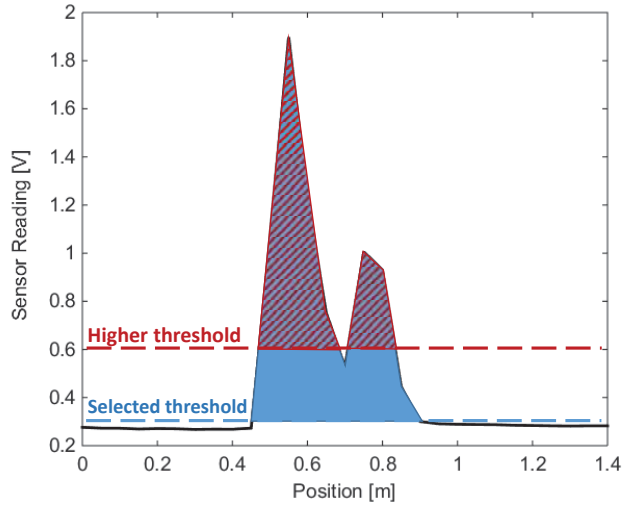
Figure 10.4 – Threshold segmentation. The number of resulting features depends on the choice of threshold level. Our selected level (blue) generates a single feature, while a higher level of 600 mV (red), would generate two.

from the reference signal, s, and the estimated signal, $\hat{s}$, denoted by their respective areas (or volumes for a 2D signal) A_1 and A_2, we define the feature localization accuracy metric as the Euclidian distance between the orthogonal projection of their centers of mass (CoMs) on the space axis (or plane for a 2D signal). An investigation of feature discrimination is beyond the purpose of this work, and consequently we selected a low threshold of 300 mV that yielded a single feature, representative of the entire plume.

■ **Root mean square error (RMSE)** - Differently from the previous metric, the RMSE quantifies the system's ability to estimate the continuous shape of the underlying signal, and is computed as:

$$RMSE = \sqrt{\frac{1}{N} \sum_k (\hat{s}(k) - s(k))^2} \tag{10.5}$$

where k represents the discrete space index, and N is the total number of samples.

The results we obtained for these two metrics are presented in Figure 10.5. The use of the Wiener deconvolution resulted in consistent performance improvement for all the considered scenarios, and for both metrics. Nevertheless the gain in RMSE quality drops significantly as the movement speed increases. We believe that this is mainly due to the lower signal-to-noise ratio (SNR) of the more damped signals measured at higher velocities. The feature localization metric does not appear to be affected to the same extent.
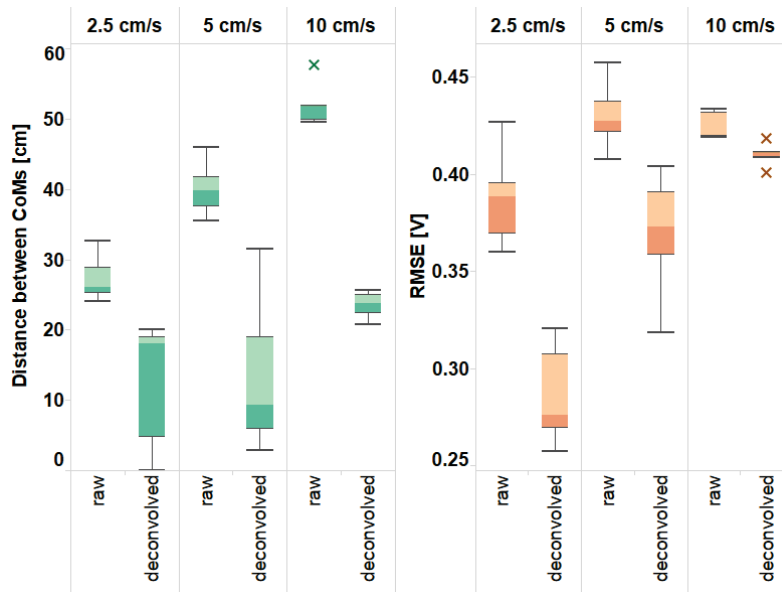
Figure 10.5 – Performance evaluation for each of the considered movement speeds: accuracy of feature localization (left), and root mean square error (right).

10.3   Air Sampling System Design

As represented in Figure 10.6, sampling system designs can be broadly classified in the following way:

- **Open passive samplers** are systems that rely only on naturally existing air flows and the relative flow induced by locomotion for transporting the target gas molecules towards the sensitive surface of the transducer. The advantage of this design rests in its simplicity and low-cost. Currently, most of the sensor nodes used in mobile WSNs applications fall in this category [33], [36], [37], [40], [42].

- **Open active samplers**, also known as sniffers, employ devices like fans or vacuum pumps to draw and flush air around the sensor for improving its dynamic response. While this approach has been used in some projects [35], [45], [67], no results analyzing the effectiveness of the considered designs has been published prior to our work.

- **Closed active samplers**, or closed chamber systems, are measurement systems in which a sample of air is drawn into a sensor chamber, usually with a pump. The chemical sensors are then exposed to this sample until their output signal stabilizes. Then the sampled air is expelled by drawing from a source of clean air. The sensor is left to recover, after which the cycle restarts. The advantage of this type of system is that, differently from the two previous classes of samplers, it can provide absolute concentration levels. However, in order to achieve the desired stability inside the chamber, a practical implementation of this measurement principle needs to consider non-trivial aspects like the reactivity of the targeted gasses to each other, and to contact
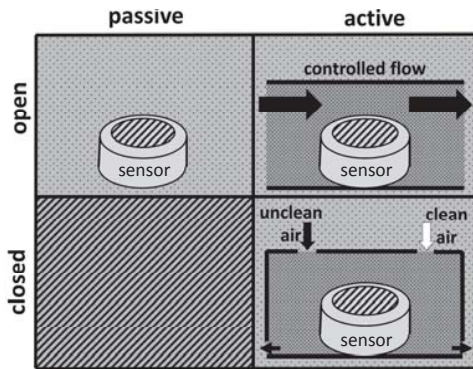
Figure 10.6 – Classification of sampling system designs. Closed passive systems do not exist, as a sensor would be isolated from external air in such a setup.
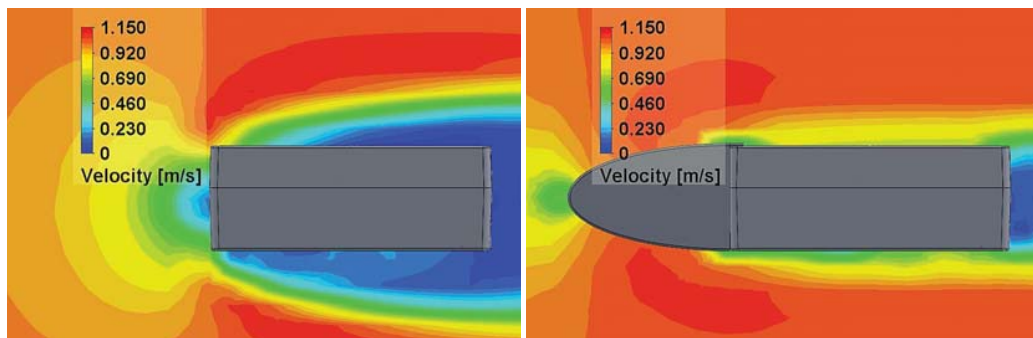
Figure 10.7 – Flow around the sensor box without (left) and with (right) aerodynamic profile.

surfaces inside the chamber and on the air-flow path. Furthermore, such a system is relatively more complicated and less light-weight, making it less attractive for mobile WSNs applications. Finally, the time needed for a complete measurement cycle implies a low measurement frequency and leads to a problem of time budgeting. To the best of our knowledge, no closed-chamber sensor node designs currently exist for mobile WSNs applications.

In our work we focused on open active samplers, which include both fan- and pump-based designs, and studied their ability to improve the measured signal in terms of SNR and response time. For the best performing sniffer, we went a step forward, by studying whether its performance is maintained when moving outside of the wind tunnel, in a real-world comparative experiment using our electric car as mobility source.

We guided our sniffer design effort by using Computational Fluid Dynamics (CFD) simulations, in order to understand the impact of our different design choices on the air flow, before moving to experimental trials. These were performed using an accurate 3D mechanical model of the sensor box and the SolidWorks Flow Simulation package.

With this approach, we first considered how the air flow around the sensor box could be improved through a purely passive solution: modifying the aerodynamic profile of the node enclosure. As can be seen in Figure 10.7, adding an elliptical profile to the front of the sensor
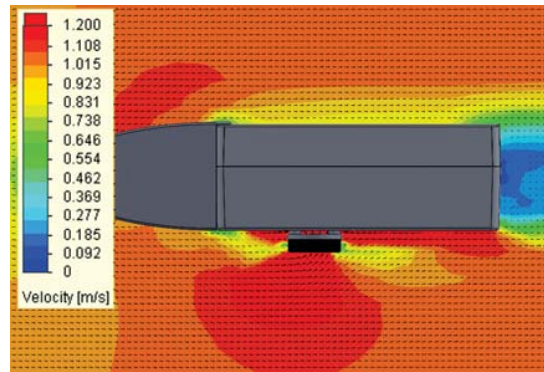
Figure 10.8 – Flow simulation for a fan-based active sampler design.

box reduces significantly the thickness of the boundary layer around the sensor box. The choice of this particular profile shape was influenced by the elliptical conical noses used in sub-sonic flying platforms. All experiments presented in this section were done using a configuration that includes this aerodynamic profile.

Beyond the simulation of external flows, the Solidworks CFD package allowed us to also simulate the effect of the actuators we considered, using models that we derived based on the performance curves supplied by the manufacturers. Figure 10.8 depicts an example of such a simulation for one of the fan-based active sampler designs we considered.

10.3.1   Fan-based sniffer designs

The fan-based sniffers we considered employ axial fans to produce air flow. Axial fans force the air to move parallel to their rotating shaft and are capable to move large volumes of air. However, their performance can be significantly influenced by the background flow conditions.

We considered two types of fan-based sniffers (see Figure 10.9):

- A *lateral flow sniffer* has the design we previously used in [48], with a fan mounted at the end of an enclosure which pulls the air over the sensor, and contains a wedge to direct flow towards the sensor.
- A *normal flow sniffer* design represents an axial fan mounted with its shaft perpendicular to the sensing surface through a supporting structure. In this case the air is pushed directly towards the sensor.

Multiple axial fans were tested on these designs with flow rates ranging from 3.74 m^3/h to 33 m^3/h.
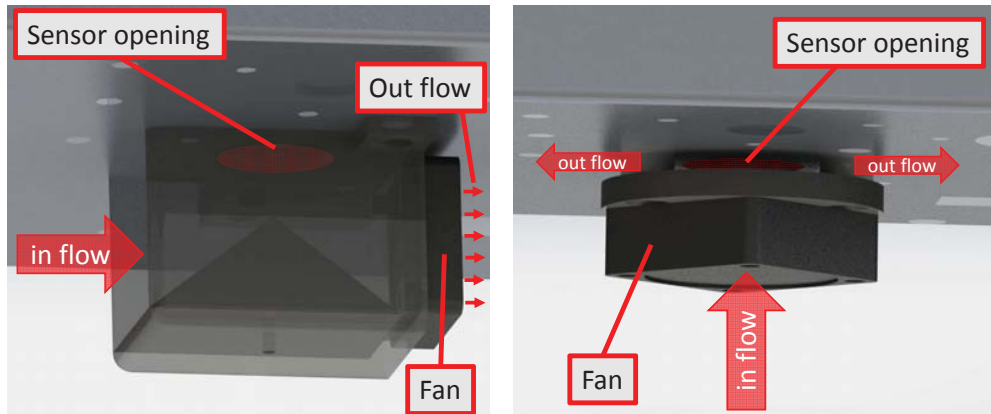
Figure 10.9 – Lateral flow (left) and normal flow (right) fan-based sniffers.
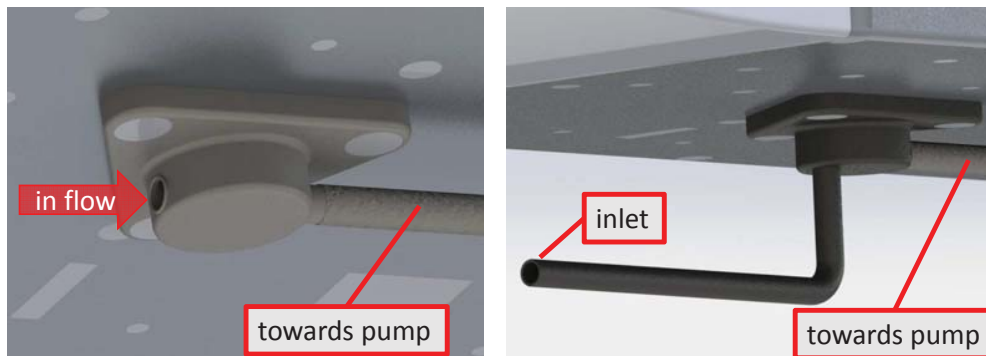
Figure 10.10 – Leveled-inlet (left) and raised-inlet (right) pump-based sniffers. In both cases the air is pulled through a narrow channel towards the sensor surface and is then evacuated through the pump out through an small opening at the back of the sensor box.

10.3.2 Pump-based sniffer designs

For the pump-based designs we used the 0.22 m^3/h diaphragm pump (Sensidyne AP240DEEE Model 60 [89]). This type of device is a positive displacement pump, having the advantage of ensuring an almost constant air flow, being less susceptible to environment conditions. Nonetheless, the volume of air that diaphragm pumps can move is small relative to their size.

In this case also, we considered two designs of active samplers (see Figure 10.10):

- A *leveled-inlet sniffer* is a basic pump-based design similar to the one used by Lochmatter et al in [86], and has the inlet opening in close proximity of the sensor box surface.
- A *raised-inlet sniffer* is a modification of the above design, in which the inlet is extended and raised outside of the boundary layer formed around the sensor box.

In order to choose the distance from the box surface that the inlet needs to be raised to, we used CFD simulations and the definition of the boundary layer thickness, as being the level
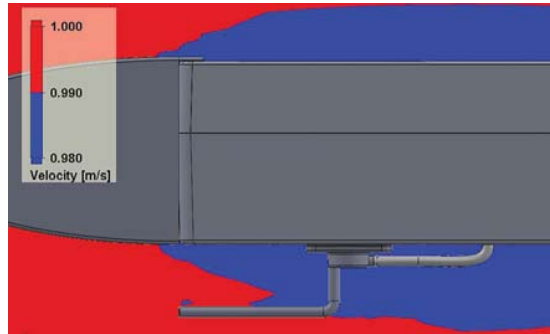
Figure 10.11 – A clearance of 3 cm from the box surface ensures that the inlet escapes the boundary layer (in blue above). The simulation in this figure was performed for a background air flow of 1 m/s.

at which the air velocity is equal to 99% of the free flow velocity (see Figure 10.11). We, thus, found that a clearance of 3 cm from the sensor box surface is sufficient to place the inlet outside of the boundary layer.

10.3.3 Results

We are interested in enhancing the measured signal from two points of view: increasing the SNR, and improving the sensor dynamic response (i.e. decreasing the sensor rise time t_r). We use the common definition of the SNR as the ratio between the power of the signal and the power of the background noise. The separation between noise and signal is done based on the assumption that the signal is null outside of the smoke plume (i.e. a direct measurement of sensor noise).

In order to estimate the sensor rise time, we again use Matlab's System Identification Toolbox to determine the parameters of a time-delayed over-damped second-order linear sensor model of the form:

$$H(\omega) = e^{-j\omega t_d s} \frac{A_0}{(1 + j\omega\tau_0)(1 + j\omega\tau_1)} \tag{10.6}$$

where A_0 is the zero-frequency gain of the model, τ_0 and τ_1 are its two time constants, and t_d is the system time-delay. This is a refinement of the sensor model described in Equation (10.4), which provides a better fit to the experimental data. Estimating this parameter is permitted by the higher frequency miniPID data, compared to our previous sparse static measurement-based ground truth estimation. In the identification process we use the miniPID data as input and the electrochemical sensor data as system output. Figure 10.12 shows an example of an identification input-output data-set pair, and the identified model output.

We performed sets of ten experiments for each sampler system configuration, traversing speed, and wind speed we considered. Throughout all our experiments both the electrochemical sensor and the miniPID were sampled at 10 Hz. We performed movements at constant speed on the Y-axis in increments starting from 5 cm/s and up to 30 cm/s (i.e. approximately 1 km/h). However, due to the narrow width of the plume (approx. 20 cm), most of the experiments were performed at the lower end of this range.
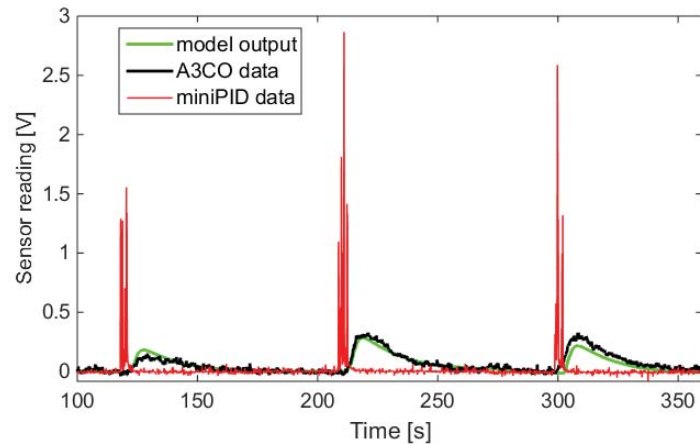
Figure 10.12 – Partial view of system identification data-set. A complete set uses 10 experiments.

Other parameters that we systematically varied are the wind speed, from 0.97 m/s up to 2.62 m/s, and the level on the Z-axis of the traversing system. We performed in total more than 800 single experiments. The results presented in this section are the most significant we obtained.

Figure 10.13 presents the performance in terms of SNR obtained with our sampling systems for different coordinates on the Z-axis of the traversing system. The traversing speed for these experiments was 5 cm/s, and the wind speed was 0.97 m/s. As can be seen, the pump-based sniffers perform better than the fan-based sniffers. The leveled-inlet sniffer performed at least as well as the passive sampler, if not better, while the raised-inlet sniffer was consistently the best of all the sampling systems considered, irrespective of the considered traversing height.

The fan-based sniffers, however, illustrate an inconsistent performance - outperforming the open passive sampler for a certain traversing height, but performing worse than it for another. It should be noted that the presented results for the fan-based sniffers consider only the configuration with the best performing actuator, which for both types of sniffers was the small 3.74 m^3/h Delta AFB02512HH [90]. This was the least powerful fan considered in our study. In fact, for the normal-flow sniffer design, we were able to observe a monotonously decreasing relationship between the performance of the sniffer and the nominal airflow of the fans we considered (see Figure 10.14). This is probably due to a mix between re-circulation of air around the fan-supporting platform, and the destructive effect of the fan on the plume structure, which we could actually observe visually.

In order to further test the robustness of the SNR enhancement by the raised-inlet pump sniffer, we performed experiments in which we incrementally varied the wind speed inside the channel, and compared the performance of the open passive sampler and the sniffer. The results in Figure 10.15 show that, while there is a general decrease in the SNR with increasing wind speeds (due to the narrowing of the smoke plume), the raised-inlet pump-based sniffer consistently outperforms the open passive sampler.
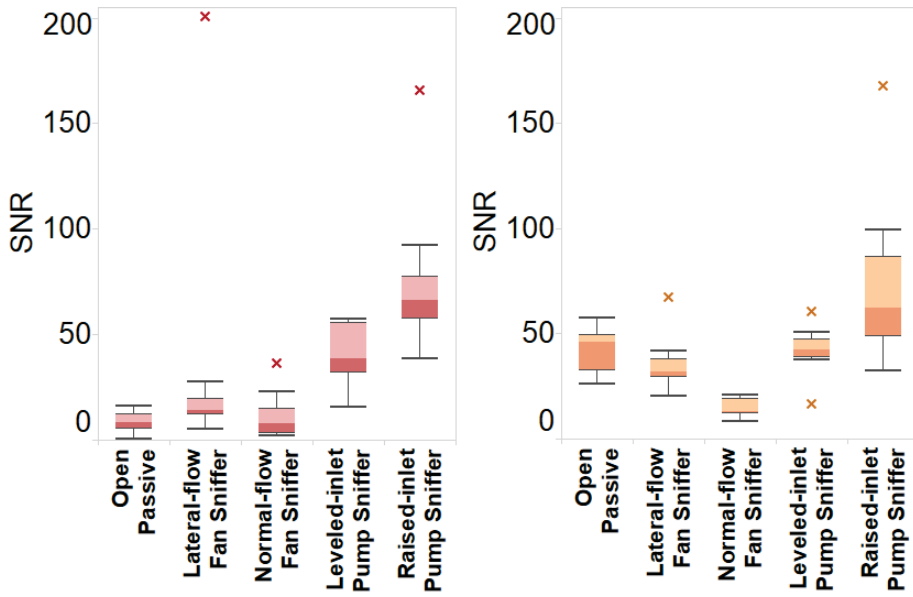
Figure 10.13 – Comparison of sampler system performance at a height of 49 cm (left), and 51 cm (right).

Figure 10.14 – The SNR of the normal flow fan-based system designs drops with the increase of the nominal fan flow. These experiments were performed at a height of 51 cm.
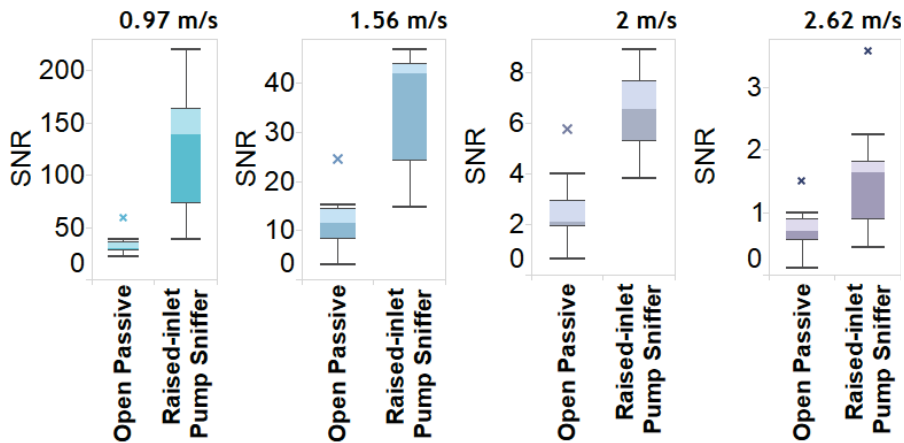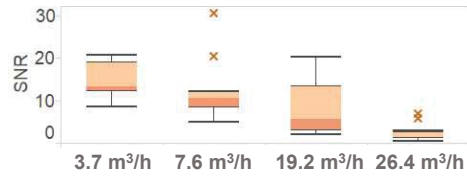
Figure 10.15 – The performance of the raised-inlet pump-based sniffer is robust to large variations of the wind speed. These experiments were performed at the same level as the smoke machine (i.e. 45 cm from the floor).
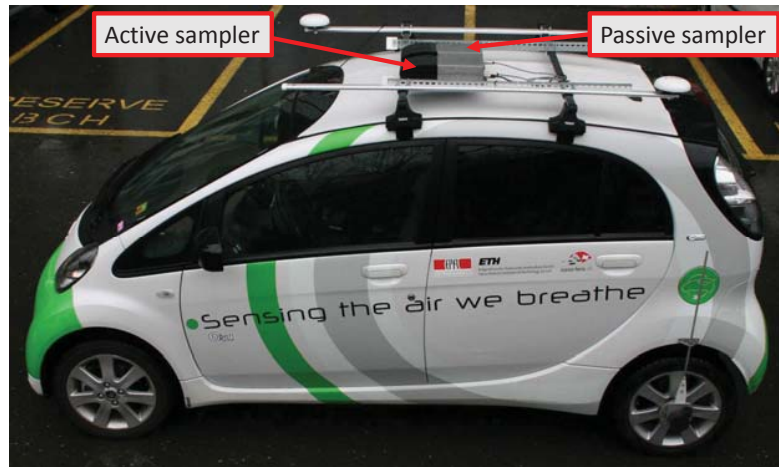
Figure 10.16 – Our electrical car equipped with a passive sampler and an active sampler in parallel.

While we were able to have an effect on the SNR with our different sampler designs, there was no observable effect on the sensor dynamics, with the estimated values for the rise time being clustered around the value of 29.2 seconds, with a standard deviation of 2.9 seconds, in line with the manufacturer's specifications (i.e. less than 40 seconds).

10.3.4 Outdoor experimental validation

After concluding that the raised-inlet pump-based sniffer provided the best performance in our controlled experiments, we decided to test it in an uncontrolled outdoors experiment. To this end, we equipped our electric car with two gas sensor boxes with similar SNR characteristics inside the wind tunnel (see Figure 10.16). One of the two sensor boxes was equipped with a raised-inlet pump-based sniffer, while the other one was left as a passive sampler.

We performed an experiment of more than one hour, on March 3, 2016, a day with moderately difficult weather conditions (wind gusts of up to 26.3 km/h and moderate sleet, during the experiment). We drove from the outskirts of the EPFL campus to the center of Lausanne where we performed ten loops in an area with regular traffic, and then returned to EPFL (see Figure 10.17).

The results of this experiment are shown in Figure 10.18. As can be seen from this graph, the two measured signals are very well correlated, but the peaks of the sensor box using our sniffer are generally higher in amplitude. The SNR values calculated for the two data-sets confirm a very robust enhancement (more than double the SNR) of the raw measurement signal provided by the active sampler.

Figure 10.17 – The trajectory of our outdoor experiment

Figure 10.18 – The results of the real-world experiment confirms the benefit of using the active sampler for enhancing the SNR.

Summary

In this chapter we investigated the impact of mobility on low-cost chemical sensor measurements. We showed that this mobility causes significant signal distortion even at low speeds. We then proposed two complementary techniques of mitigating this effect. The first one is through deconvolution, a signal processing technique that tries to estimate the underlying signal based on a sensor model and information on the measurement noise. Using a Wiener filter deconvolution we managed to obtain significant signal improvement in terms of both feature localization accuracy and RMSE. The effectiveness of deconvolution is, however, limited by the typical poor SNR of the measured signal, which is exacerbated with increasing platform speed. In order to enhance the SNR of the measurements, we proposed a second technique: the use of an active sampler. We considered multiple sniffer designs and found that their use can indeed improve the SNR of the chemical sensor measurements.

11 Conclusion

DEVELOPING the appropriate techniques for mitigating the effect of slow-dynamics of low-cost chemical sensors is essential for the real-world success of mobile WSNs for air quality monitoring. In this part of the thesis we experimentally analyzed the effect of mobility on readings of a chemical sensor in a controlled wind tunnel environment. We showed that mobility has a significant effect on the chemical sensor readings even at the low movement speeds we considered, and would certainly be very severe for more realistic speeds.

To increase the quality of measurements we proposed and studied two different, but complementary solutions: (i) a deconvolution technique that tries to reconstruct the underlying signal, and (ii) the use of an active sampling system to enhance the quality of the raw measured signal. In this chapter we discuss the performance of both of these approaches critically, outlining possible directions for improvement.

11.1   Signal Deconvolution

Although the Wiener deconvolution consistently produced positive results, there is still room for improvement. One direction that would be useful to investigate is the further refinement of the assumed sensor model. One option is the use of a two phase model, with different dynamical properties for the rise and recovery behavior of the sensor, which have been reported to be asymmetrical in previous works [85].

In our work we used the data measured at different sensor movement speeds to derive a single LTI model for the distortion of the signal. This is an appropriate approach as long as the distortion is caused exclusively (or predominantly) by the sensor dynamics in conjunction with the platform kinematics. At urban vehicle speeds, we expect that flow perturbation induced by the platform mobility will play a larger role. In this context, a single linear distortion model, independent from platform speed, will probably not be sufficient, and might need to be replaced by either a piece-wise linear model or a nonlinear model.

11.2   Active Sampling

We investigated the potential of active samplers to enhance the quality of chemical sensor readings in mobile applications for air quality monitoring. We performed rigorous repeatable experiments in a controlled environment - a wind tunnel - to study the benefit of using sniffers for this type of application.

We considered both fan- and pump-based sniffer designs and found that the latter are generally better suited as a sniffer actuator. In particular, one of our proposed pump-based designs, which places the inlet opening outside of the boundary layer formed around the sensor box, delivered consistently superior results in SNR enhancement compared to the baseline passive sampler. This result was confirmed also outside the controlled environment of the wind tunnel, through a real-world vehicle-based experiment that we conducted using two sensor boxes in parallel.

However, no significant improvement of sensor dynamics was achieved with the considered sniffers. This might be due to the different sensor technology of the sensor we considered, compared to the ones used in works that have reported the opposite [85] (i.e. electrochemical vs. semiconductor).

We conclude that mobile sensing systems for air quality monitoring can indeed benefit, at least in terms of SNR enhancement, from the use of well designed active samplers. That being said, the design of the sniffers we considered are not optimized. Although they are based on extensive simulation and experimental simulation trials, further effort needs to be placed into the systematic study of their design.

Finally, we want to emphasize that, for maximizing signal reconstruction quality, active sampling and deconvolution should be used in conjunction, as the increase in SNR provided by the former technique will enhance the performance of the latter.

Summary

We have proven that both signal deconvolution and the use of active sampling are valid techniques for mitigating the distorting effect of mobility on chemical sensor measurements. This is however still an open research direction, as both techniques should be further improved. The hypothesis of a single LTI system for the sensor model needs to be relaxed, as we expect it to be insufficient at higher platform speeds, as air flow perturbation start to play a bigger role. Also, the design process of active samplers needs to be approached in itself in a more systematic way to enhance the benefit of their use.

Calibration Part IV

12 Introduction

T HE most widely used chemical sensors for WSNs applications are either electro-chemical or semiconductor. Both of these classes of sensors suffer from temporal drift, different degrees of cross-sensitivity, and temperature, humidity or pressure dependence. This lack of stability and selectivity poses a tough challenge for ensuring basic data quality and needs to be addressed through adaptive sensor calibration techniques.

12.1 Sensor Calibration

Ensuring sensor calibration is essential for environmental monitoring systems, since it allows to relate sensor readings to the physical quantity of interest. Factory calibration of low-cost chemical sensors is usually not provided. In the rare cases when it is available, the value of this calibration can be low, depending on the difference between the calibration and deployment conditions. The actual in-field performance of the sensors greatly depends on specific electronic integration and the application domain, the latter determining the range of variation in environmental conditions, but also the degree of exposure to other chemicals to which they might be sensitive. As such, the task of ensuring sensor calibration typically needs to be addressed by the system integrator.

The different options for performing calibration can be broadly classified as *offline*, or *online* methods. The first class includes one-off initial calibration methods, by experimental evaluation with a parallel reference instrument before deployment [91], [92]. The success of this type of methods depends on the ability of the experimental set-up to re-create the variability of all the relevant influencing parameters. Although costly and not trivial, this might be feasible for characterizing cross-sensitivities and the effect of environmental parameters. However, accurately capturing temporal drift behavior can imply impractically long experiments. An option for dealing with this issue would be replacing the one-off calibration scheme with a periodic offline calibration, but when considering large scale deployments this would be too time-consuming to be practical.

Online calibration algorithms are automatic methods for estimating the sensor calibration parameters during deployment. To state the problem, consider a set of low-cost mobile sensor nodes, each having its own specific characteristics in terms of temporal drift, cross-sensitivity to other chemicals, and environmental parameter dependencies. The question is how to calibrate them in order to have consistently accurate measurements. Mobile WSNs provide an opportunity for performing this type of calibration by leveraging spatiotemporal encounters between pairs of sensor nodes, called *rendezvous*, and using the measurements gathered within these events to improve the calibration parameter estimate.

If no reference signal is available, this type of methods can provide a relative calibration of the network (i.e. ensuring consistency between sensor nodes). If rendezvous with a reference station are available, these methods can also provide an absolute calibration. In this chapter, we assume that there is at least one reference sensing station which the mobile sensors encounter from time to time. Using the measurements taken by the sensors and the reference station, a statistical regression method can potentially provide a calibration function.

The opportunity for rendezvous calibration that mobile systems offer comes, however, at a price. As discussed in Part III, mobility induces a distorting effect on chemical sensor measurements, due to the typical long response time of this kind of sensors. This effect is not negligible, and needs to be considered when discussing rendezvous calibration.

12.2 Previous Work

Pioneering contributions to the field of online rendezvous calibration are the works of Hasenfratz et al. [78] and Saukh et al. [79], [80]. In [78], the authors introduce and compare three types of online algorithms (forward, backward, and instant calibration) based on least squares temporal-weighted regression of calibration tuples (i.e. pairs of reference and raw sensor measurements). They evaluate the performance of these algorithms by considering both simulated and real measurements, with uniformly distributed simulated rendezvous events. In [79], Saukh et al. give a rigorous definition for rendezvous events and introduce the concept of rendezvous connection graphs, studying their properties in the context of sensor fault detection and sensor calibration. More recently, in [80], the authors address the problem of error accumulation in multi-hop calibration algorithms that use ordinary least squares regression, and propose the use of geometric mean regression as a better alternative to the former method.

Both [79] and [80] consider a linear relationship between sensor readings and the targeted concentration level, using only two parameters: the sensor's baseline and its gain. Based on this assumption, the proposed algorithms attempt to estimate these parameters directly, and assume them to be close to constant between calibration cycles. The sensor model considered in [78] is a generalized polynomial relating sensor readings to gas concentrations. The effect of environmental parameters or the temporal drift are not included in the model, although the latter is taken into account indirectly by the weighing mechanism of the proposed calibration
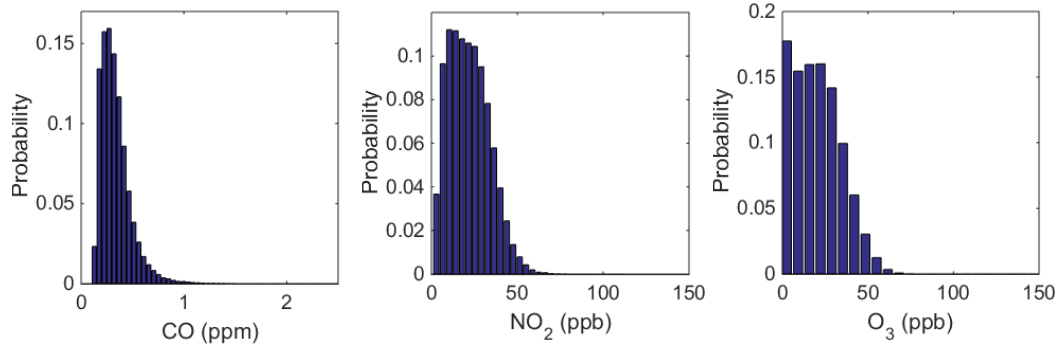
Figure 12.1 – Distribution of CO, NO_2 and O_3 concentrations measured by Lausanne NABEL station since the start of our deployment. Note that the scale of CO is in ppm, since it is a gas that is much more present in the considered environment than either NO_2 or O_3.

algorithms.

None of the previous works have considered the effect of mobility on chemical sensor measurements, using instead the assumption of point measurements (i.e. instantaneous in space and time).

12.3 Pre-deployment Offline Calibration

For the sake of completeness, before discussing rendezvous calibration and our contributions on this front, we present a study of static offline calibration using one of the sensor nodes we have deployed at the NABEL station site. We consider a hypothetical scenario of pre-deployment calibration, whereby we would statically calibrate each sensor node for a certain time period, using the NABEL station as reference, before installing them on a bus.

The calibration time period we consider is of one week. We calibrate each of the chemical sensors in our node, except for CO_2, which, as we mentioned in Chapter 5, has an automatic self-calibration algorithm. We derive each calibration function starting from the manufacturer specifications. To help understand the quality of the resulting calibration, Figure 12.1 shows the distributions of pollutant concentrations reported by the NABEL station for CO, NO_2 and O_3.

For the CO sensor, the datasheet information [57] claims high selectivity and linearity with respect to the targeted pollutant. The auxiliary electrode of the sensor has the role of compensating for environmental parameters variations (e.g., temperature, humidity, etc.). Although slow temporal drift is mentioned, no characteristic for this behavior is provided. Taking all this into consideration, we consider a basic linear calibration function for this sensor:

$$CO^{cal} = p_0 + p_1 \cdot CO^{raw} \tag{12.1}$$
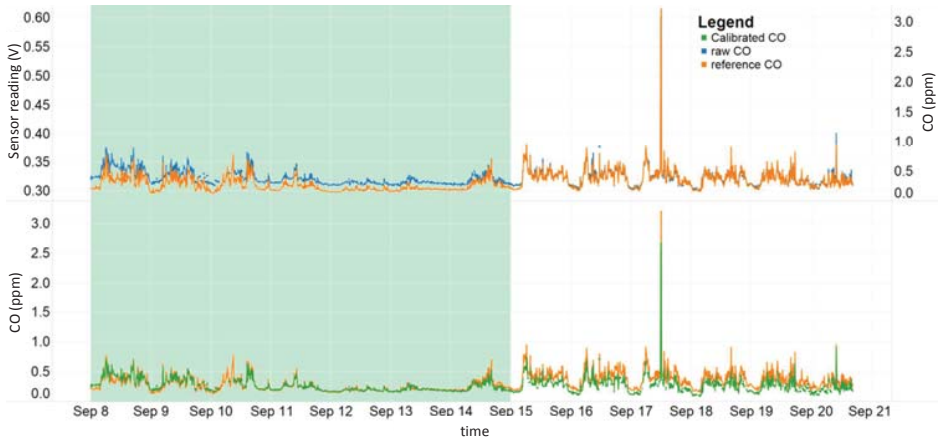
Figure 12.2 – CO sensor measurements before (top) and after (bottom) calibration. The calibration period is highlighted in green. The drift of the calibrated series is visible in the second week.
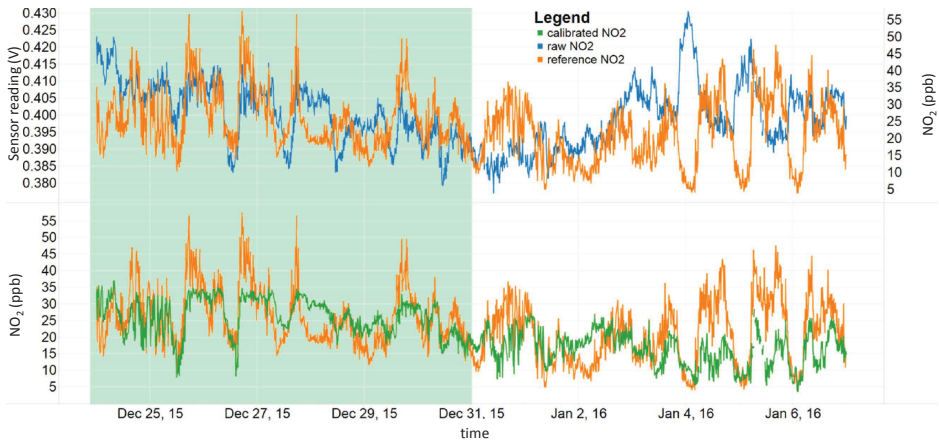
Figure 12.3 – NO_2 sensor measurements before (top) and after (bottom) calibration. The calibration period is highlighted in green.
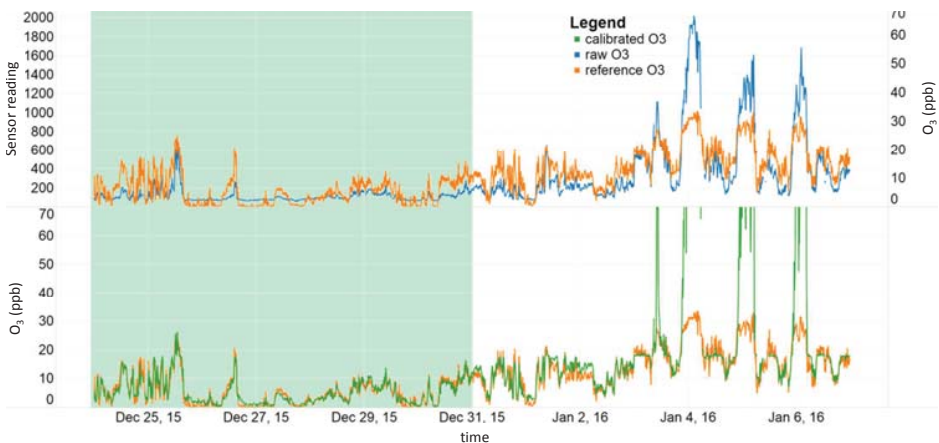
Figure 12.4 – O_3 sensor measurements before (top) and after (bottom) calibration. The calibration period is highlighted in green. Because of the high non-linearity of the sensor, and most-likely an over-fitted model, the error already reaches very high levels in the second week.
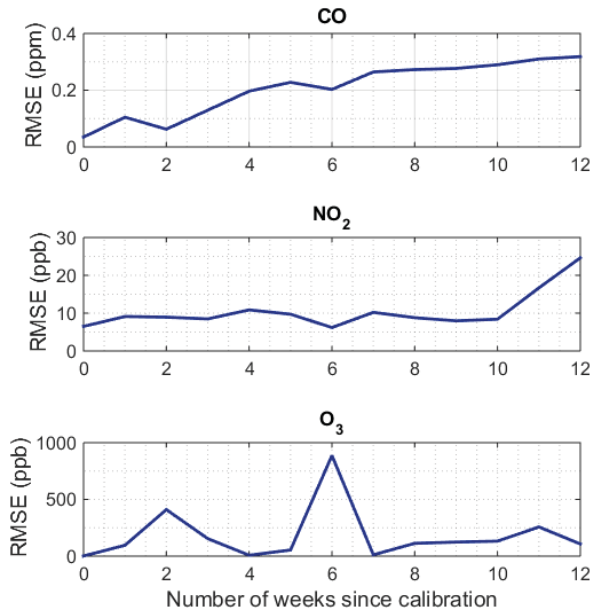
Figure 12.5 – Evolution of calibration error (RMSE) calculated per week. Note that while the error for the electrochemical sensors (CO and NO_2) increases at a relatively steady pace, the error evolution of the highly non-linear O_3 semiconductor sensor is much more erratic.

The available information on the NO_2 sensor is very limited [58]. What we do know is that it is cross-sensitive to multiple other species. One of these is O_3, and, as such, we consider a linear combination between the sensor reading and the NABEL O_3 reference. We opt to use the NABEL data, instead of our own sensor in order to not confound the sources of calibration error. The calibration function we assumed for this sensor has the following form:

$$NO_2^{cal} = p_0 + p_1 \cdot NO_2^{raw} + p_2 \cdot O_3^{ref} \tag{12.2}$$

The specifications of the O_3 are significantly richer [55]. They include information on temperature dependency, the non-linearity of the sensor response, and suggest a calibration function, which we also adopted:

$$O_3^{cal} = p_0 + p_1 \cdot O_3^{t.c.} + p_2 \cdot (O_3^{t.c.})^2 + p_3 \cdot (O_3^{t.c.})^3 \tag{12.3}$$

with:

$$O_3^{t.c} = O_3^{raw} \cdot \exp(K \cdot (T - 25°)) \tag{12.4}$$

where K is a temperature dependency parameter, and T is the measured temperature.

Using these functions we performed the calibration for each sensor using one week of data. Figures 12.2 - 12.4 illustrate the obtained results for each sensor. For CO, we obtained an RMSE of 0.037 ppm during the calibration week, and a coefficient of determination (R^2) of 0.88. The results for the NO_2 sensor where relatively poor (RMSE of 6.5 ppb and $R^2 = 0.41$), however,

using ozone as calibration input, did improve the performance as a linear model that did not include it obtained an R^2 of only 0.16. Finally, the performance for the O_3 sensor within the calibration week was good, with an RMSE of 1.5 ppb, and an R^2 of 0.89.

While this performance can be considered good with respect to typical concentration ranges shown in Figure 12.1, or at least reasonable in the case of the NO_2 sensor, the quality of the calibration is lost in time for all sensors (see Figure 12.5), motivating the need for an online re-calibration approach.

Summary

In this chapter we introduced the problem of calibrating networks of low-cost chemical sensors. The available options for performing this operation can be broadly classified as either offline or online methods. We discussed these types of calibration, showing that while offline calibration methods are impractical for long-term large-scale sensor network deployments, mobile WSN deployments offer the opportunity for automatic online calibration, by exploiting inter-node rendezvous events. We presented previous work on automatic online calibration of low-cost chemical sensor networks and their underlying assumptions. Finally, we illustrated the shortcomings of pre-deployment offline calibration using static measurement data from our deployment.

13 Proposed Algorithms

ACHIEVING reliable long-term sensor calibration is an essential subject of research for the future success of air quality monitoring systems that employ low-cost chemical sensors. In the context of mobile WSNs, rendezvous events between uncalibrated nodes and calibrated nodes can be exploited to enable automatic online calibration throughout the deployment lifetime. While, in recent years, significant progress has been made in developing such calibration techniques, this remains an open research topic.

In this chapter, we present our own contribution to this research effort. We start by investigating the opportunity of increasing the complexity of the sensor models considered in rendezvous calibration algorithms. We follow this by a relaxation of the point measurement assumption used in previous works, by tacking into account mobility-induced measurement distortion.

13.1 Model-based Rendezvous Calibration

As discussed in Chapter 12, previous works on online rendezvous calibration have considered only a simple linear relationship between sensor readings and the level of the targeted pollutant (i.e. a linear sensor model). We propose a different approach to the calibration problem by placing a direct focus on the sensor model selection and its impact on calibration performance.

In particular, we consider models of varying complexity and a sliding-window model-based rendezvous calibration algorithm and evaluate its performance for different sensor models and different assumptions on the calibration window size.

13.1.1   Sensor model

The general form for a model describing the behavior of a chemical sensor can be expressed as:

$$s(t) = f(t, y(t), \mathbf{x}(t), \mathbf{w}(t), e(t)) \tag{13.1}$$

where:

- t represents the time variable
- $s(t)$ is the sensor reading
- $y(t)$ is the true concentration value of the targeted chemical
- $\mathbf{x}(t)$ is the vector of chemical signals to which the sensor is cross-sensitive
- $\mathbf{w}(t)$ is the vector of environmental parameters (e.g., temperature, relative humidity, etc.) which have an effect on the sensor
- $e(t)$ is a zero-mean independent noise.

As most chemical sensors are designed to have a linear dependence on $y(t)$, we introduce the baseline and gain functions $b(\cdot)$, $g(\cdot)$, and re-write Eq. 13.1 as:

$$s(t) = b(t, \mathbf{w}(t)) + g(t, \mathbf{w}(t)) \cdot y(t) + f_x(t, \mathbf{x}(t), \mathbf{w}(t)) + e(t) \tag{13.2}$$

In particular, if the sensor has a good selectivity for the targeted chemical, then $f_x(\cdot)$ can be dropped, leading to:

$$s(t) = b(t, \mathbf{w}(t)) + g(t, \mathbf{w}(t)) \cdot y(t) + e(t) \tag{13.3}$$

In this case, calibrating the sensor means obtaining the estimates $\overline{b}(\cdot)$, $\overline{g}(\cdot)$ of the sensor baseline and gain, in order to infer the targeted chemical concentration:

$$\overline{y}(t) = (s(t) - \overline{b}(t, \mathbf{w}(t)))/\overline{g}(t, \mathbf{w}(t)) \tag{13.4}$$

In previous work on rendezvous calibration, the gain and bias functions were considered to be constant between calibration cycles [78], [80]. The significance of this assumption on calibration performance depends on the inter-play between the degree of sensitivity to environmental factors, drift dynamics, and the frequency of rendezvous events with a reference station.

To investigate this, we continued to focus on the City Technology A3CO carbon monoxide sensor. The A3CO is a highly selective 4-electrode electrochemical sensor, which means that the general model structure in Eq. 13.3 can be assumed. As is the case with other electrochemical sensors, temperature variations can affect both the sensor gain and its baseline. Despite the fact that the sensor's forth electrode is meant to compensate for temperature effects on the baseline, its use does not completely eliminate temperature

influence. The sensor is also sensitive to pressure variations. While this might be a problem when using pump-driven sampling, for the passive sampling system considered in this work, this was not an issue. Finally, the sensor manufacturer states that there is no significant effect of humidity variation on the sensor performance, which was confirmed by our experiments.

To derive a list of the candidate sensor models, we have used data from sensor nodes deployed for long periods at the NABEL station in Lausanne (see Fig. 7.3), whose time series are used for ground truth purpose. Apart from the A3CO readings and the NABEL reference for CO, we also record with our node the temperature and relative humidity. Based on the sensor specifications and our analysis of the collected time series, we propose the following candidate sensor models of increasing complexity, which include explicit consideration of temporal drift and dependence on temperature (denoted as $T(t)$):

$$m_1 : s(t) = p_0 + p_1 \cdot y(t) \tag{13.5}$$

$$m_2 : s(t) = p_0 + p_1 \cdot y(t) + p_2 \cdot t \tag{13.6}$$

$$m_3 : s(t) = p_0 + p_1 \cdot y(t) + p_2 \cdot t + p_3 \cdot T(t) \tag{13.7}$$

$$m_4 : s(t) = p_0 + (p_1 + p_2 \cdot T(t)) \cdot y(t) + p_3 \cdot t + p_4 \cdot T(t) \tag{13.8}$$

$$m_5 : s(t) = p_0 + p_1 \cdot y(t) + \frac{p_2}{p_3 + t} + p_4 \cdot T(t) \tag{13.9}$$

$$m_6 : s(t) = p_0 + (p_1 + p_2 \cdot T(t)) \cdot y(t) + \frac{p_3}{p_4 + t} + p_5 \cdot T(t) \tag{13.10}$$

Before investigating a calibration algorithm, we first tested the quality of each of these models by fitting them through least squares regression on the full static data-sets gathered from our nodes deployed near the NABEL station (which was used as the reference). We evaluated the performance of the models through a 10-fold cross-validation technique. The model fitting was done both on the full time range, but also on decreasing temporal sections (i.e. monthly, weekly, daily). The performance metric used was the Root Mean Square Error (RMSE) of the calibrated series.

An exemplar result is presented in Figure 13.1 for the sensor which was the longest continuously deployed at NABEL (approximately 10 months). As can be seen, for the complete time range, the more complex models outperformed the simpler ones, although adding a temperature effect to the gain in models m_4 and m_6 did not produce a significant improvement, and they were subsequently discarded.

As the temporal calibration window is reduced, the benefit of the more complex models is gradually lost, and while for the monthly and weekly scales having at least an explicit temporal component still brings a marginal benefit, this disappears completely at the daily scale. This performance trend is justifiable by the fact that, at shorter time scales, the temperature excursion is also limited, making it more difficult to accurately estimate its impact.
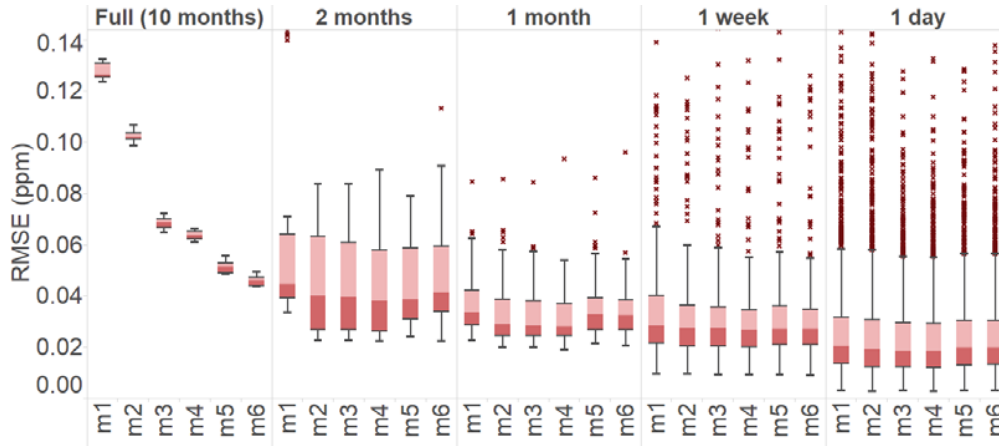
Figure 13.1 – Results of fitting the 6 sensor models on different time scales, using sensor data gathered over 10 months near the reference NABEL station. 90% of all available data is used for fitting the models, and 10% for validation.

Looking at the temporal drift, an interesting aspect is the fact that, at the monthly scale, the hyperbolic models perform worse than the linear ones. This is most probably due to the fact that a hyperbola is harder to estimate than a line when reducing the number of regression samples.

13.1.2 Rendezvous calibration

Rendezvous calibration is the technique of achieving network calibration by using data tuples gathered during rendezvous events, together with a statistical modeling method (e.g., regression analysis). A broad definition of a rendezvous is any encounter between pairs of sensors which takes place within a particular space and time bound. The underlying assumption of rendezvous calibration is that data-set pairs of measurements taken during sensor encounters will be relatively well correlated.

In previous works [79], [80], a data-set of measurements gathered by two sensor nodes i and j, with respective measurement series $\mathbf{s_i} = \{s_i\}$ and $\mathbf{s_j} = \{s_j\}$, during rendezvous was defined as:

$$\Phi = \{(s_i, s_j) \mid (|l_i - l_j| < \Delta d) \wedge (|t_i - t_j| < \Delta t)\} \tag{13.11}$$

where l_i and l_j are the positions of the two sensor nodes, t_i and t_j are the time-stamps of their measurements, and Δd and Δt are the spatial and temporal constraints that define the rendezvous. This is a straightforward and general definition of rendezvous data. However, some of its implications merit a brief discussion.

Adjusting the two constraining parameters, Δd and Δt, allows a tuning of the trade-off between the likelihood of a good correlation between the rendezvous data-sets and their size. Both of these aspects directly impact the quality of the model fitting.
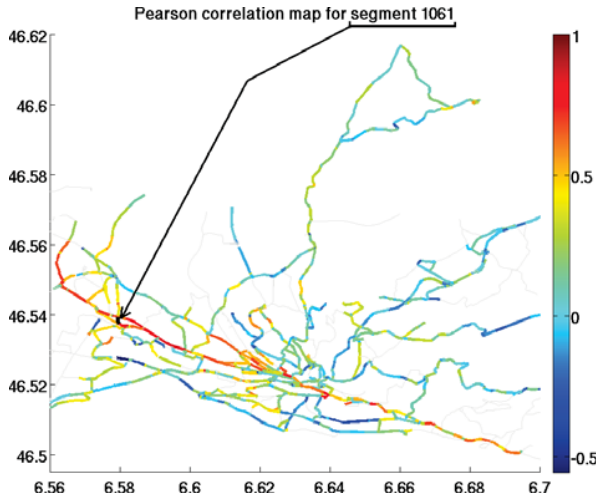
Figure 13.2 – The correlation map for segment 1061 (shown by a small black dot and the arrow), for LDSA measurements. The segments along one long street show high correlation while some nearby streets in other directions do not show high correlations.

The merit of using the Euclidean distance between sensor nodes for deciding whether or not they are within rendezvous range is debatable for an urban scenario. In fact, in an urban environment measurement correlations are based on traffic flows through the street network more than on the geographic distance. Figure 13.2 illustrates this in the case of our particulate matter measurements.

In our work, we used a more conservative and restrictive definition for a rendezvous event: two stations are considered to have had a rendezvous only if they are present at the same time on the same street link (i.e. on the same segment of a street, between two junctions). This transforms the definition in Equation (13.11) to:

$$\Phi = \{(s_i, s_j) \mid (\{l_i, l_j\} \subset \Psi_n) \wedge (t_i = t_j), n = \overline{1, N}\} \tag{13.12}$$

where Ψ_n is the set of all possible sensor node positions within a given street-link n, out of the complete N-length street network.

The calibration method we propose is illustrated in Algorithm 1. Considering i the calibrated node, and j the uncalibrated node, the routine is executed at a given time (t_{curr}). It assumes a particular sensor model (m), and a calibration time window (w). All the data pairs that satisfy the rendezvous condition are collected into a calibration data-set (S_{cal}), including the temporal and environmental parameter information for the uncalibrated node (t_j and w_j, respectively). If the correlation between the pairs of data-sets is significant (p-value < 5) and larger than a threshold (th_{corr}), and the number of data tuples is sufficient for fitting the particular model, a least squares regression is performed. This yields a set of sensor model parameters, which can then be used to compute the calibration parameters - the output of the routine. For all results presented in this chapter the value of th_{corr} was set to 0.4.

The calibration routine can be run in a single step, between a reference station and each uncalibrated node (single-hop calibration), or in multiple steps, with newly calibrated nodes helping to calibrate other nodes (multi-hop calibration). In our work we considered only

---

Algorithm 1 `Model_Based_Rendezvous_Calibration`$(\mathbf{s_i}, \mathbf{s_j}, \mathbf{w_j}, t_{curr}, w, m, th_{corr})$

1: **for all** $(t_i > t_{curr} - w) \wedge (t_j > t_{curr} - w)$ **do**
2:    **if** $(s_i, s_j) \in \Phi$ **then**
3:       `add_to_calibration_set`$(S_{cal}, s_i, s_j, t_j, w_j)$
4:    **end if**
5: **end for**
6: $pass = $ `test_correlation`(S_{cal}, th_{corr})
7: $N_p = $ `number_of_parameters`(m)
8: **if** $(pass = \text{TRUE}) \wedge (card(S_{cal}) \geq N_p)$ **then**
9:    $p_{model} = $ `fit_model`(S_{cal}, m)
10:   $p_{cal} = $ `compute_calibration_parameters`(p_{model}, S_{cal})
11:   **return** p_{cal}
12: **else**
13:   **return** NULL
14: **end if**

---

single-hop calibration using the static NABEL station as reference.

Our calibration method is a refinement of the basic forward calibration algorithm described in [78], which also uses rendezvous data tuples gathered within a certain time window to obtain the calibration parameters at a given time. The main distinctive feature comes from the fact that instead of weighing measurement tuples based on their age and regressing directly for the calibration parameters, our algorithm first estimates the sensor model parameters.

This allows us to consider more complex sensor models, which explicitly take into account temporal and environmental information. It also implies that between calibration cycles, the bias and gain are not necessarily kept constant. If a model that includes a temporal or an environmental parameter component is used, this additional information will be used to estimate the corresponding calibration parameters at a particular moment in time.

13.1.3 Performance

To evaluate the performance of the proposed calibration method and its sensitivity to sensor model choice and calibration window size, we consider two test scenarios. The first one is based on realistically simulated rendezvous, and the second one uses actual rendezvous events between our mobile nodes and the NABEL site. Figure 13.3 shows the location of the NABEL station and the adjacent street link within which a rendezvous with the reference station can take place.

Simulated rendezvous

We devise the following set-up: we use one of the static sensor nodes deployed at the NABEL station as a mock mobile station and we simulate rendezvous events based on realistic
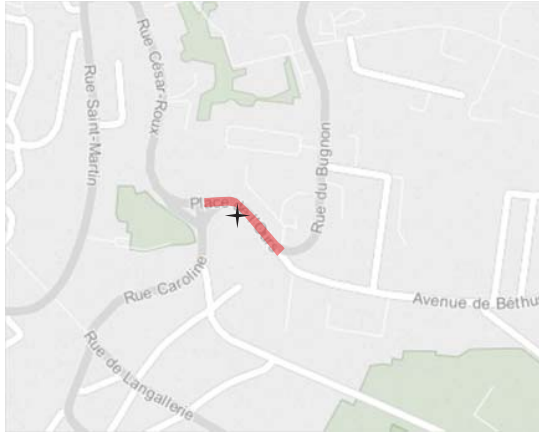
Figure 13.3 – The location of the NABEL station is marked with a black star and the corresponding rendezvous street link is highlighted in red. The length of the link is of approximately 120 m.
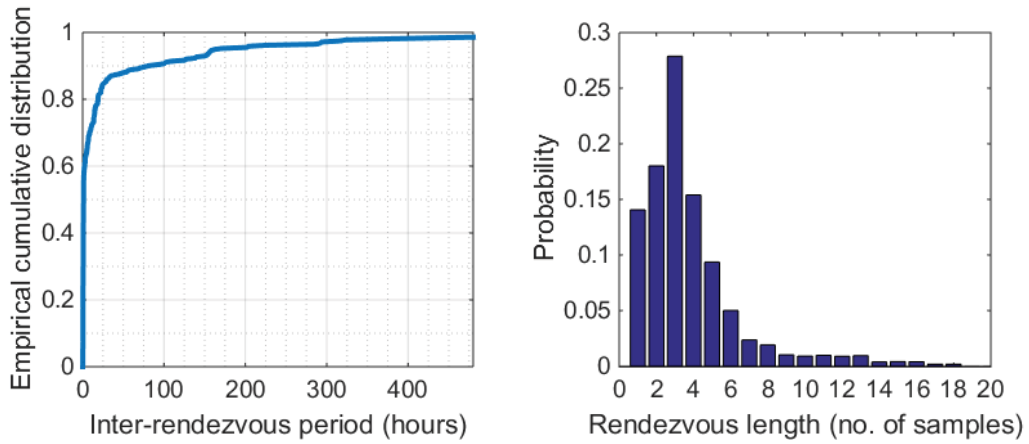
Figure 13.4 – Empirical cumulative distribution function for inter-rendezvous periods (left) and probability mass function of rendezvous length (right) are derived from our mobile sensor network.

probability functions. The advantage of this approach is that the ground truth is known throughout the experiment, allowing us to calculate the performance of the algorithm also between calibration cycles. Furthermore, this set-up allows us to isolate the performance evaluation of the algorithm from the effects of mobility on measurement quality, which are not accounted for at this stage.

Unlike the scenario described in [78], the NABEL station in Lausanne is not located on a regular bus route, and the rendezvous with it are therefore sparse and irregular. To derive realistic probability functions for sampling our simulated events, we have used the mobility data generated by our bus-anchored nodes (see Figure 7.1) during their first 22 months of deployment. Considering only rendezvous events between one bus and the reference station, we calculated the empirical cumulative distributed function of the inter-rendezvous period and also the probability mass function of the length of one rendezvous in number of datapoints at the default sampling frequency of 0.2 Hz (see Figure 13.4). Based on these functions we simulate realistic rendezvous sequences for our deployment.
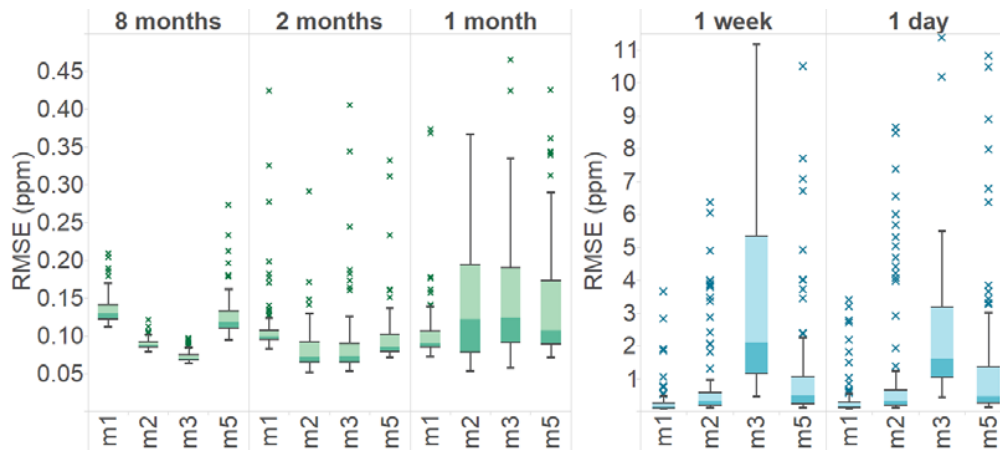
Figure 13.5 – Evaluation of model-based rendezvous calibration algorithm for the four selected models using calibration windows of different sizes. The performance drops, particularly for short calibration windows because of the rendezvous distribution.

We ran the calibration algorithm, varying the calibration window from one day to 8 months and using all models except m_4 and m_6. For each combination of the set-up parameters we performed 100 randomized runs. The results of these experiments are shown in Figure 13.5.

Due to the scarcity of data, for the short calibration windows the complex algorithms perform poorly, but as the window increases the situation is reversed. In fact, the performance obtained by complex algorithms with wide calibration windows is superior to the one obtained by the simplest model on short windows.

The main drawback of using the complex models is the necessary size of the window in itself, which implies that the system needs to wait for a long time before being able to do the first calibration. One practical solution to this problem is, of course, to use both types of models: the less complex ones during the early phase of the deployment and the more complex one as it matures.

These results are relevant for a scenario of sparse rendezvous events and the performance of the algorithm will tend to the regression results presented in Figure 13.1, as rendezvous become more frequent.

Real-world rendezvous

The next natural step was to apply the proposed calibration algorithm for mobile measurement data based on the real-world rendezvous. Differently from the simulated rendezvous case, when using mobile measurements we do not have a complete ground-truth data series and need to rely exclusively on the rendezvous data for both running the calibration and evaluating its performance.
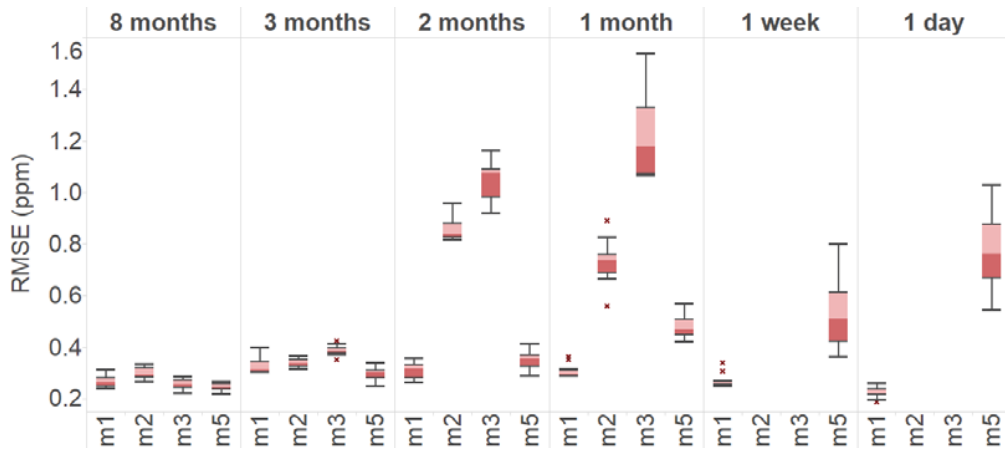
Figure 13.6 – Performance evaluation of the rendezvous calibration algorithm for the sensor node deployed on bus 602. The error of models m_2 and m_3 when using the weekly and daily calibration windows are very high and not visible on this graph, which was clipped on the Y-axis, for clarity.

In order to evaluate the calibration performance, we proceeded to split the rendezvous data pairs for each bus-anchored node into training and validation sets through a randomized 10-fold cross-validation technique. We then applied the calibration algorithm for each mobile node using the same settings as for the simulated rendezvous (i.e. the same type of models, window sizes and correlation threshold).

An exemplar result set for one of the mobile sensor nodes is shown in Figure 13.6. While the actual individual calibration performance for each node varies as a function of the number and distribution of rendezvous, this graph is illustrative for the overall behavior of the proposed algorithm.

As was the case for the simulated rendezvous, the more complex sensor models perform more poorly when using shorter calibration windows, but their performance relative to the simplest linear model (m_1) gets better as the window is increased. However, differently from the simulation case, the best performance obtained when using complex models (for the 8 months window) does not appear to be significantly better than the best performance of model m_1, and actually, for the particular case in Figure 13.6, it is clearly worse.

Finally, the performance over all types of models and calibration windows drops compared to the simulation results. This is not surprising. The measurement pairs in this case are no longer taken closely together, and as such are less correlated. Also, we are expecting the mobile measurements to suffer from distortion because of the long response time of the chemical sensors.

13.2   Mobility-aware Rendezvous Calibration

As we showed in Part III, the slow dynamic response of chemical sensors induces significant measurement distortion when used on a mobile platform. If left unaddressed, this issue will certainly contribute to the reduction in performance quality of rendezvous calibration algorithms. In this section, we investigate the opportunity of using the deconvolution technique, presented in chapter 10, for improving the performance of such algorithms.

We propose modifying the calibration routine presented in the previous section by adding to it a preprocessing step in which we apply a deconvolution using a Wiener filter. The resulting procedure is illustrated in Algorithm 2. The dynamic sensor model we use for the deconvolution is the same as in Equation (10.6), a time-delayed over-damped second-order linear system.

Algorithm 2 `Mobility_Aware_Rendezvous_Calibration`($\mathbf{m_{dyn}}, \mathbf{s_i}, \mathbf{s_j}, \mathbf{w_j}, t_{curr}, w, m, th_{corr}$)

1: $\mathbf{s_j^f} \leftarrow$ `deconvolve`$(\mathbf{s_j}, \mathbf{m_{dyn}})$
2: **for all** $(t_i > t_{curr} - w) \wedge (t_j > t_{curr} - w)$ **do**
3:    **if** $(s_i, s_j^f) \in \Phi$ **then**
4:       `add_to_calibration_set`$(S_{cal}, s_i, s_j^f, t_j, w_j)$
5:    **end if**
6: **end for**
7: $pass =$ `test_correlation`(S_{cal}, th_{corr})
8: $N_p =$ `number_of_parameters`(m)
9: **if** $(pass = \text{TRUE}) \wedge (card(S_{cal}) \geq N_p)$ **then**
10:   $p_{model} =$ `fit_model`(S_{cal}, m)
11:   $p_{cal} =$ `compute_calibration_parameters`(p_{model}, S_{cal})
12:   **return** p_{cal}
13: **else**
14:   **return** NULL
15: **end if**

We performed the calibration of all the mobile nodes across all the different scenarios described in the previous section, and using the same randomized 10-fold cross-validation technique. Due to the technical difficulty of dismounting all the CO sensors in our network and performing the wind tunnel experimental trials needed for individually identifying their time constants, we decided to work under the assumption of a low variability of the response dynamics across sensor cells. As such, we used the same dynamic sensor model parameters for deconvolving the measurement signals from all the mobile nodes.

The results of using the Wiener deconvolution for the calibration algorithm yielded predominantly positive results. Figure 13.7 shows an exemplar set of results obtained with our mobility-aware algorithm compared to the basic rendezvous calibration algorithm presented in the previous section.
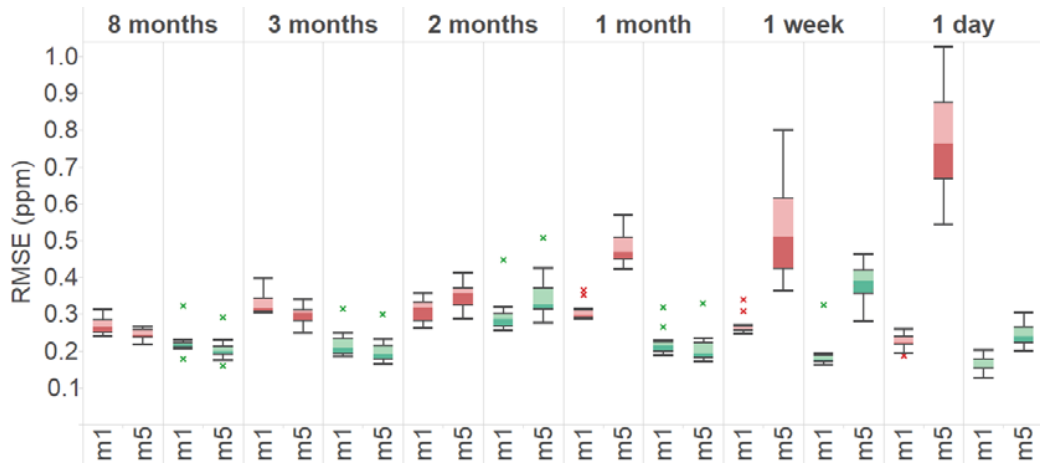
Figure 13.7 – Comparison of calibration performance when not using the Wiener deconvolution (red), and when using our mobility-aware approach (green) for the sensor node on bus 602. For visual clarity, only the results obtained when using models m_1 and m_5 are presented.
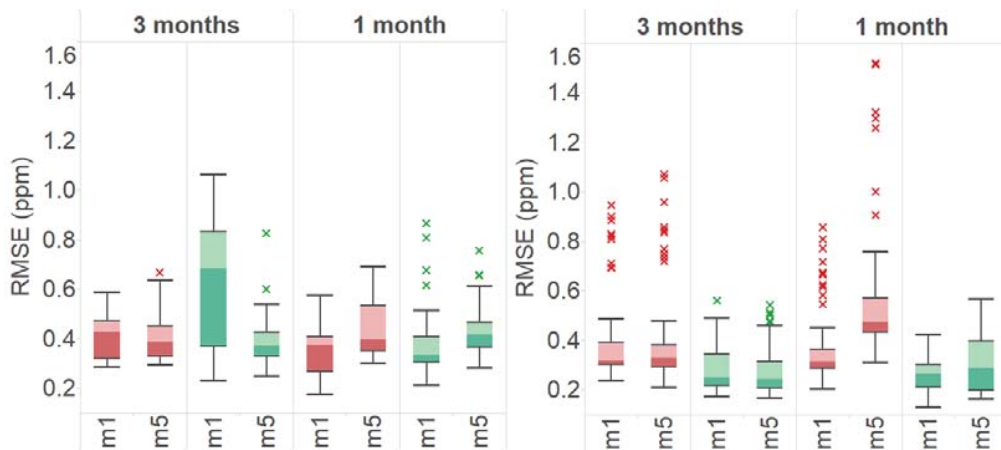
Figure 13.8 – Aggregated results for buses 606 - 608 (left) show mixed or worse performance when using deconvolution (green), than when not using it (red). Aggregated results for the rest of the deployment are shown for reference (right).

The use of the signal deconvolution consistently reduced the calibration error obtained across the different calibration window sizes, and sensor models considered, for most of the mobile sensor nodes in our network. However, the degree of improvement varied across sensors nodes, and for three of them the resulting performance was either not improved or even reduced (see Figure 13.8). These relatively mixed results point to a probable inadequacy of the assumption on low variability of the response dynamics across sensors.

Summary

In this chapter, we investigated two new directions for improving on the state-of-the-art in rendezvous calibration algorithms for mobile sensor networks employing low-cost chemical sensors. First, we proposed a model-based approach to the calibration problem. We analyzed the ability of sensor models with increasing levels of complexity to explain the behavior of the CO electrochemical sensor used in our deployment. These included models that explicitly take into account temporal drift and temperature effects. We subsequently proposed and implemented a sliding window rendezvous calibration algorithm based on the selected models. We evaluated its performance for different models and varying calibration window sizes, using both simulated and real-world rendezvous events. Finally, we studied the benefit of adding a deconvolution preprocessing step to the algorithm, in order to mitigate the distorting effect of mobility on chemical sensor measurements.

14 Conclusion

E NSURING a sufficient degree of sensor calibration for the targeted application represents a fundamental data quality requirement for any sensor network deployment. When considering low-cost chemical sensors, solving this task is particularly challenging because of their typical instability and cross-sensitivity.

In our work on this subject, we proposed two new directions for pushing further the performance of rendezvous calibration algorithms: (i) by considering sensor models that are more faithful to their complex behavior, and (ii) by accounting for the distorting effect of mobility on sensor measurements. In this chapter we present a critical discussion of the obtained performance for each of these approaches, placing our contribution in the context of related work, and outlining possible directions for further improvement.

14.1 Model-Based Approach

Based on our study of the long-term behavior of the electrochemical sensors used in our deployment, we proposed a model-based approach to online rendezvous calibration of mobile air quality monitoring networks. To this end, we derived models with increasing levels of complexity, including models that explicitly take into account temporal drift and temperature effects.

We then evaluated the ability of each of these models to approximate the sensor behavior at different temporal scales and found that the more complex models are more accurate when considering relatively large temporal scales. As the temporal scale is reduced, the benefit of the more complex models is gradually lost, and while for the monthly and weekly scales having at least an explicit temporal component still brings a marginal benefit, this disappears completely at the daily scale.

Using a selection of the most promising sensor models, we proposed and implemented a sliding window rendezvous calibration algorithm, evaluating its performance for varying set-

up parameters (i.e. model type, and calibration window sizes). We first considered a simulation set-up that used realistically modeled rendezvous sequences and real measurement data from a static station. There are at least two benefits to this approach: First, it allowed us to evaluate the performance of the calibration algorithm throughout the whole time range (including the time between calibration cycle). Second, it provided us an insight into the limit of the calibration quality attainable for the considered rendezvous conditions (i.e. sparse and non-uniform), by discarding the negative effects of mobility.

Our simulation results showed a potential for calibration performance increase when using more complex models over larger calibration windows, compared to basic linear models over short time windows. Nevertheless, when assessing the performance of the same algorithm for mobile measurement data (i.e. real rendezvous), this benefit was no longer visible, as the overall calibration quality dropped, and became more variable.

It should be noted, however, that the A3CO electrochemical cell we focused on is a relatively stable and selective sensor. Sensors with higher degrees of deterministic temporal drift and cross-sensitivities will most probably benefit more from the use of more complex models than a basic single-variable linear one. In the future, we intend to expand on this work by considering the other sensors in our deployment: the cross-sensitive electrochemical NO_2 sensor, and the considerably less stable semiconductor O_3 sensor.

Based on our results so far, we consider that the complexity of the model should be carefully chosen as a function of both sensor characteristics and the temporal distribution of calibration opportunities. While in this work our approach was heuristic, we are interested in the development of statistical methods that would be able to tune the model complexity in order to maximize calibration performance and to reduce the sensor characterization effort.

We are the first to consider the question of complex sensor models in the context of rendezvous calibration of low-cost chemical sensors. In a very recent work, Maag et al. [81] have also considered the benefit of using more complex sensor models for improving calibration. In particular, their work focused on mitigating chemical sensors cross-sensitivities. However, their work proposed a pre-deployment, offline derivation of calibration parameters, and did not investigate the possibility of estimating them online. Moreover, the possibility of modeling the sensor drift was not investigated.

14.2 Mobility-Aware Approach

Building upon our work on mitigating mobility effects on chemical sensor response, we proposed a mobility-aware approach to rendezvous calibration algorithms. As such, we augmented our calibration algorithm with a preprocessing step in which the measurement signal of the mobile node was first passed through a Wiener deconvolution filter. We assumed the dynamic response model with the same parameters that we identified based on our wind tunnel experiments, presented in Part III.

While our results indicated that our assumption of very low variability of dynamic response parameters across sensors was probably flawed, they also showed the potential for enhancing calibration quality that a mobility-aware calibration approach has. However, in order for this technique to be useful, the problem of correctly identifying the time constants for each sensor cell needs to be solved. As in the case of estimating the steady state sensor model parameters, two different approaches to this problem can be considered: offline and online.

While still time consuming, the pre-deployment identification of the dynamic response of each sensor is more feasible than for estimating steady-state parameters, since the experiments would require only fast variations of the targeted gas, and not the exploration over the full range of all the external factors that the sensor is susceptible to. Nevertheless, an automatic online approach would be preferable also in this case, and an interesting concept would be that of estimating the dynamic model parameters together with the other sensor model parameters through the use of rendezvous data. However, achieving good results with such an approach is difficult to imagine in the context of sparse, irregular rendezvous events and one-hop calibration.

This brings as to another important point: while in our work to date we only considered one-hop calibration (i.e. direct calibration between a mobile node and a static reference station), a natural further step would be the implementation of a multi-hop schema. The advantage of the multi-hop approach has already been shown in the work of Saukh et al. [80]. In this case of rendezvous between two mobile nodes, we expect an even higher benefit of a mobility-aware calibration approach.

Finally, as we are in a scenario of relatively sparse physical meetings between the mobile nodes and the reference station, the way we define a rendezvous should be investigated more closely. A relaxation of our very strict rendezvous definition to perhaps include spatially more distant street segments, which are nevertheless well correlated, might provide the necessary data for further improving the calibration quality.

Summary

We conclude our work on chemical sensor calibration for mobile networks by discussing critically the performance of our two approaches. We place our contribution in the context of related work and discuss the directions for further improvement. While our work on both fronts is very novel and we showed the benefit of using both model-based and mobility-aware rendezvous calibration, there is still room for improvement. More specifically, the use of model-based approaches needs to be applied to different types of chemical sensors, and the tuning of sensor complexity should be automated. Also, in the case of the mobility-aware approach, the problem of correct estimation of dynamic model parameters needs to be solved. For both approaches, the impact of implementing a multi-hop calibration strategy should be investigated.

High-Resolution Air Pollution Part V Mapping

15 Introduction

ASSESSING the relationship between human health and long-term exposure to urban air pollutants is a major goal of many medical studies. This is a critical problem considering the fact that millions of premature human deaths are annually linked to air pollution [1]. In the framework of numerous medical projects (e.g., PIAMA [93], CoLaus [94] and SKIPOGH [95]), various health related parameters (e.g., blood pressure, renal salt excretion and physical activities) and location of a large number of participants are frequently collected. For physicians, one of the main missing pieces of the puzzle to accurately link specific medical disorders with air pollution are high spatio-temporal resolution pollution maps of the urban areas in which participants reside and work.

While the use of mobile sensor networks increases the spatial resolution of measurements, this cannot on its own, deliver complete air pollution maps. This is certainly true for our own mobile deployment, as its spatial coverage in respect to the urban area of interest is in any case partial: the public transportation network is limited to only a subset of streets, meaning that we never have any measurement for the streets/areas that are not covered. Moreover, the coverage changes over time because of the dynamic assignment of buses, based on the needs of the public transportation company. Therefore at any given period of time there are many parts of the city that do not have any measurements.

Given these limitations and the high spatial heterogeneity of pollutants in urban areas, generating consistent and complete maps with high spatio-temporal resolution is a tough challenge which needs to be addressed through air pollution modeling.

15.1   Air Pollution Modeling

Physics-based simulation and *statistical modeling* are the two main categories of modeling frameworks for air pollution estimation [96]. Using emission inventories and meteorological data as input, physics-based simulations use physico-chemical models of airborne gas dispersion to estimate pollutant concentrations. GRAL [21] is an advanced example of this category

which mathematically models the motion of pollution plume particles in the atmosphere using a Lagrangian dispersion model. One drawback of this category is that they rely on the use of accurate data on emissions (e.g., type and number of vehicles in the streets), structural and geographical details of the environment (e.g., building dimensions), and meteorological data (e.g., temperature and wind speed), which are not always available in high temporal resolutions. Moreover, these simulations cannot adapt to real-world changes respective to their set-up parameters (e.g., topography or traffic flow modifications) without manual intervention, and cannot capture particular pollution events (e.g., construction sites, fires, sudden changes in the weather conditions, etc.).

Statistical models treat other modalities correlated to a specific pollutant (e.g., co-pollutants, temperature, etc.) as *explanatory variables*. Explanatory variables are random variables helping to derive a statistical description of the target distribution. These methods can be divided into two classes. The first class is represented by the purely measurement-driven models which aim at finding all the dependencies and variables from the measurements of the targeted pollutant. Spatial interpolation methodologies (e.g., inverse distance weighting interpolation [97], and K-Nearest Neighbor (KNN) [98]) are the most common approaches in this class. The performance of such methods drops drastically if the pollution distribution is dynamic and multi-variant (which is usually the case for urban environments under short term observational conditions). The second class consists of statistical models which work not only based on the field measurements but also take one or more explanatory variables into account. These methods usually show higher performance compared to the purely measurement-driven models.

As concrete examples of the second class of statistical models, Niska et al. [99] designed an architecture of a Multi-Layer Perceptron (MLP) Artificial Neural Network (ANN) model for forecasting hourly concentrations of NO_2 at an urban station. They proposed a parallel genetic algorithm for selecting the input variables. Yi et al. [100] proposed another ANN structure for modeling ozone based on meteorological variables in an urban environment through a pattern recognition approach. Hussein et al. [101] developed a linear regression model on the data of a stationary monitoring node to predict aerosol particles. Mølgaard et al. [102] proposed a Bayesian regression model to predict UFP concentrations at an urban monitoring station using meteorology and traffic data as inputs. To obtain better prediction performance, Clifford et al. [103] proposed a generalized additive model using meteorological data, time, solar radiation and rainfall as explanatory variables. Reggente et al. [104] employed a Gaussian process regression to estimate UFPs in an urban air pollution monitoring network based on local and remote concentrations of NO_x, O_3 , CO, and UFPs. None of the mentioned works have considered mobile sensor networks as source of in-situ measured data.

One thrust of research has focused on modeling the air pollution based on land-use data. Land-use features (in the context of environmental engineering) are measures of population density, building heights, heating type, terrain elevation, terrain slope, types of roads, and so on, and even sometimes average traffic volumes. Li et al. [77] proposed a Gaussian process

regression (also called Kriging) model using land-use characteristics to estimate urban UFP levels on a grid-based spatial partitioning, using measurements collected from urban trams.

The main problem with land-use data is that usually they are not available in high temporal resolutions, implying that, they are usually representative for trends in air pollution (i.e. on a larger time scale). One way to overcome this issue is to generate different models for every targeted time period. Hasenfratz et al. [67] and Li et al. [77] (in two separate works) built up two sets of a thousand models, each targeting one time period (e.g., one model per day) for one city. These models cannot be used for time periods other than those they have been trained for. In the method proposed by Hasenfratz et al. [67], measurements gathered in a previous period are also used in the model to increase the accuracy of high-resolution maps. In particular, they annotate the UFP measurements obtained during one year with the corresponding meteorological (e.g., temperature) and time data (e.g., weekday). Then based on the current meteorological conditions and time, they fetch the most relevant historic UFP measurements and use them to augment the current dataset represented by the real-time UFP measurements. This method significantly increases the accuracy of the maps, although the meteorological data time series are not directly used in the model itself. On the other hand, Li et al. [77] did not consider meteorological parameters at all.

15.2 Building a Statistical Modeling Framework

In our work, we investigated statistical modeling for estimating air pollution maps at different temporal scales. We proposed five different statistical modeling methods, with varying degrees of complexity, using measurement data gathered by our mobile sensor network deployment.

We focused in particular on the PM sensor modality (i.e. LDSA measurements). The reasons for doing so are twofold. First, particulate matter is one of the most interesting types of pollutants we are measuring from a health impact perspective. Second, the electrical detection instruments employed in our network do not suffer from the same drawbacks as the chemical sensors (i.e. they are fast-responding, relatively stable, and factory calibrated), and, as such, the problem of developing pollution mapping techniques could be isolated from potential artifacts of the pre-processing steps required for chemical sensor measurements (e.g., residual errors in signal reconstruction or calibration). That being said, if an appropriate handling of these chemical sensor-specific issues is assumed, the statistical pollution modeling techniques presented in this part are generalizable also to the other modalities monitored by our sensor network.

The data set used for studying all of the proposed modeling techniques is identical and spans over a period of more than 14 months, starting from the inclusion of the PM detectors in our network, in December 2013. This represents more than 44 million LDSA data-points. This amount of data is available after data cleaning (i.e. applying several simple filters based on the device health meta-information).

Figure 15.1 – Left: One snapshot of the developed street matching software. In this example the blue lines show the reported trajectory traces of the buses during one week in one segment. Due to localization errors, the blue traces are deviated form the actual position of the street segment. Our street matching algorithm has matched all of them to one segment. The green lines show the other segments that buses have passed during this period. Right: The histogram of length of the street segments. Most of our segments are shorter than 25 meters.

In this section, we introduce the basic building blocks of our statistical modeling framework, by discussing the choices of spatio-temporal domain discretization and explanatory variables.

15.2.1 Domain discretization

We consider "quarterly", "monthly", "weekly", and "daily" time resolutions. Depending on these, the LDSA measurements and the explanatory variables containing temporal information were partitioned and aggregated in time, producing a list of *aggregated data points*.

Most of the previous works (e.g., [67], [105]) partitioned the area of interest into uniform grid cells and assumed that the measurements inside a cell have the same conditions (e.g., in terms of weather, wind and traffic). However, depending on its size, one cell can cover several streets which can potentially have different environmental and traffic conditions. To overcome this issue, Jutzeler et al. [106] proposed to use regions of homologous emissions to divide the city into partitions with similar daily traffic estimations. They associated every measurement to the closest road segment based on Euclidean distance. They showed that compared to grid-based, the region-based partitioning produces better predictions across aggregates of yearly to daily time scales. We follow this concept while using a more advanced street matching algorithm.

We assumed a street map representation of Lausanne based on the OpenStreetMap collaborative project database [107]. We then proceeded to a street segmentation procedure, which split street segments at junctions, and, furthermore, divided very long streets into multiple smaller streets (no longer than 600 m) in order to not lose too much of the spatial resolution. The use of this space discretization naturally resulted into higher resolutions in the downtown areas where street segments are shorter and the heterogeneity of the measured field is expected to

also be higher than in suburban areas. Figure 15.1 shows the length histogram of the street segments.

Using the localization data of the measurements and estimating the azimuth of the bus, the LDSA values are assigned to their corresponding road segments based on the algorithm explained in [108]. The general idea of this algorithm is to continuously track the buses based on their location and to keep a list of route candidates for them during their movements in the streets. Each route has a score that defines how well the traced trajectory of the bus matches this route. Figure 15.1 shows one snapshot of our street matching software and shows an example of how well the measurements are assigned to one street segment.

All the LDSA measurements were projected on 1337 street segments covering the region of interest depicted in Figure 15.2. As the figure shows, the measurements are unevenly distributed in various segments, representing a dynamic non-uniform coverage.

An important metric of the goodness of a particular spatial discretization is how homogeneous are the measurements inside a given partition element (i.e., segment or grid cell), with better discretizations having lower deviation from the mean of the partition element. We have compared the standard deviation of LDSA values in our street segments with grid-cell partitioning considering six different cell numbers in Figure 15.3. This figure shows that the standard deviation we obtain from street segmentation with 1337 segments is better then when we use 4900 (and for some cases 10000) grid cells. These results show the benefit of "street segmentation" over "grid partitioning" in urban environments.

15.2.2 Explanatory variables

In our work, we considered explanatory variables of different types. These include proxy pollutant concentrations, meteorological, land-use, and traffic data.

For the first two types of variables we considered time series (without a spatial dimension) from static monitoring stations. These data are available as 10 min averages for each of the considered parameters, and as such represent sets of about 70,000 rows each. The land-use data, on the other hand, is a set of spatial explanatory variables, with no temporal component. Finally, the traffic data we used represents a predominantly spatial explanatory data set, but includes also some limited temporal information.

Proxy pollutant concentrations

The source of our proxy pollutant explanatory data is the NABEL monitoring station in Lausanne. We selected the following pollutant species monitored at this site:

- **CO**, **NO** and **NO$_2$**: These gases are mainly produced by combustion of fossil fuels and so they can be a good measure for traffic conditions in the city. Since PM includes an
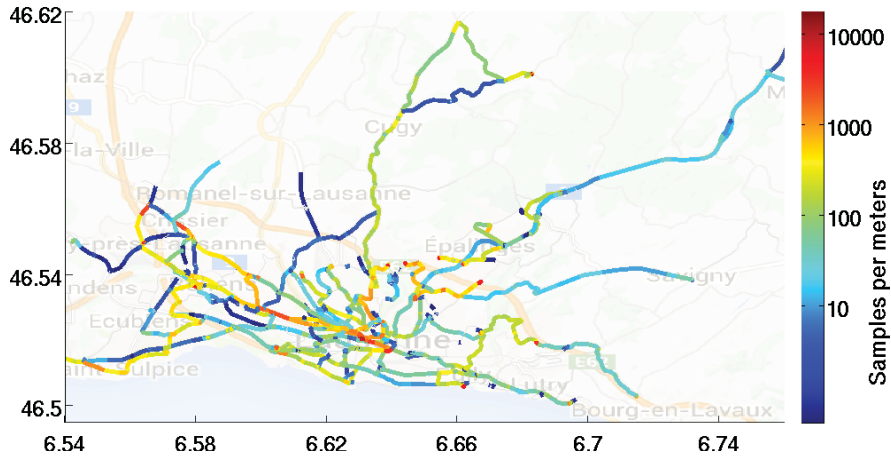
Figure 15.2 – The coverage area of the sensors from Dec. 1st. 2013 to Feb. 1st 2015. The number of samples is normalized considering the length of each link. The segments with less than 1 sample per meter are not considered in the coverage area.
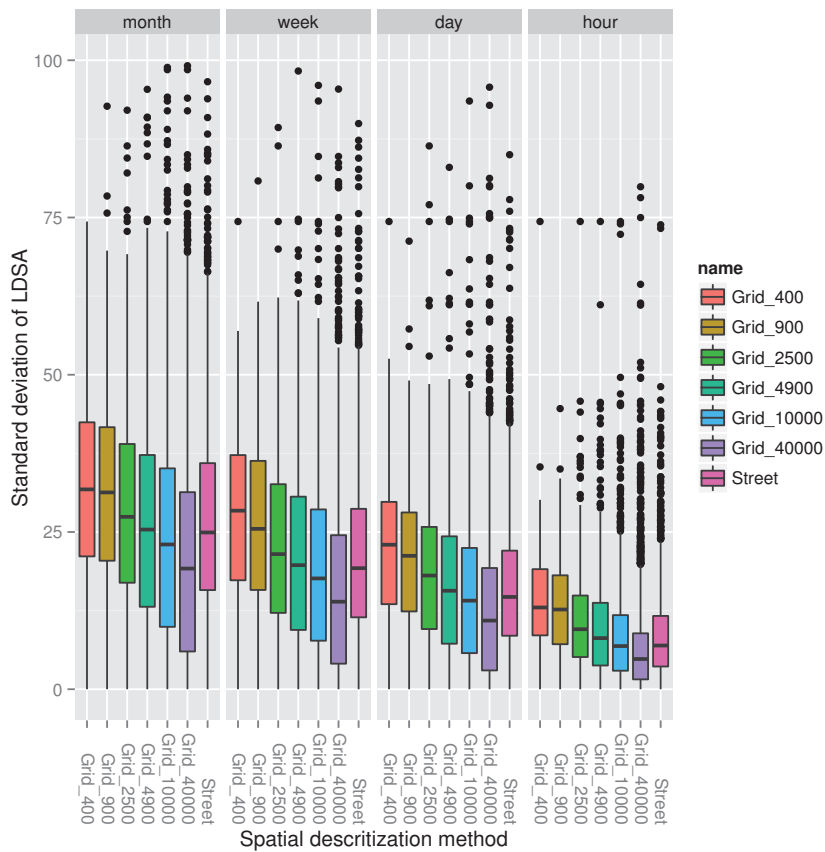
Figure 15.3 – Standard deviation of LDSA values in one spatial partition (cell or street segment). Lower standard deviation indicates more homogeneity in the measurements in one cell. The measurements are integrated based on four time resolutions and then the standard deviation in each cell/segment is computed for every space-time tessellation. Street segmentation shows good results considering that the number of segments is 1337.

120

Table 15.1 – Selected modalities measured by the NABEL and MeteoSwiss static stations (left). The parameters of land-use data obtained from the Federal Office of Topography and the Federal Office of Statistics (right).

Parameter	Unit
CO	mg/m^3
NO	$\mu g/m^3$
O_3	$\mu g/m^3$
NO_2	$\mu g/m^3$
Rain	mm
Radiation	W/m^2
Temperature	K
Relative Humidity	$\%$
Wind speed	m/s

Parameter	Unit
Altitude	m
Slope	angle in $°$
Population	#/ha
Households	#/ha
# of buildings	#/ha
# of buildings using gas heating	#/ha
# of buildings using oil heating	#/ha
Industry	#/ha
Primary industries	#/ha
Secondary industries	#/ha
Tertiary industries	#/ha

important primary component generated by traffic, these pollutants are good candidate proxy parameters.

- **Ground level O_3**: While its relationship with particulate matter is unclear, ozone is the primary oxidant of pollutant gases present in the atmosphere. It plays an important role in the balance between NO and NO_2 in the atmosphere, and species oxidized by O_3 can represent precursors for the formation of secondary PM.

Meteorological data

The meteorological data used for our models came from the MeteoSwiss station in Pully, an eastern suburb of Lausanne. The retained parameters from this station were:

- **Radiation**, **precipitations**, **temperature** and **wind speed**: The stability of the atmosphere is highly dependent on these parameters. Solar radiation and temperature change the size of eddies which eventually affect the concentration of particles through dispersion. When a precipitation event (e.g., rain) starts, the concentration of the particles drastically drops, because of wet deposition. Also wind is generally expected to disperse locally the aerosols from one place to another. It is therefore important to take these parameters into account.
- **Relative humidity**: The growth pattern of particles is related to adsorption of water vapor, so humidity is a relevant parameter to model aerosols.

Statistical analysis of the correlation between these parameters and PM measurement data has previously been studied in the literature (e.g., [109]).

The full list of the proxy pollutant and meteorological data from the NABEL and MeteoSwiss stations used in our modeling framework is summarized in Table 15.1.
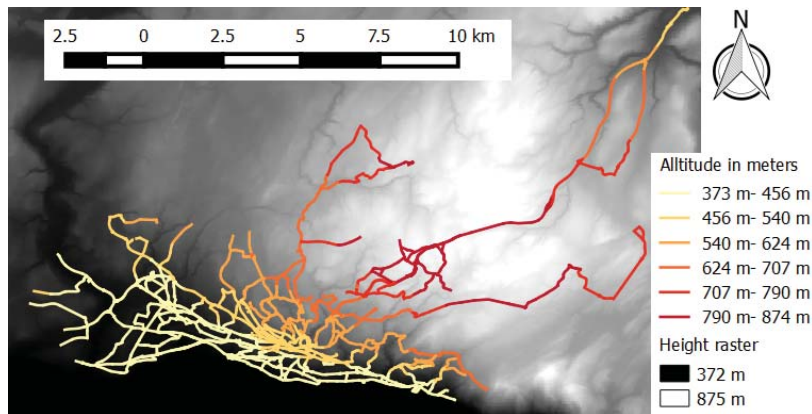
Figure 15.4 – Altitude profile of street segments of the city.
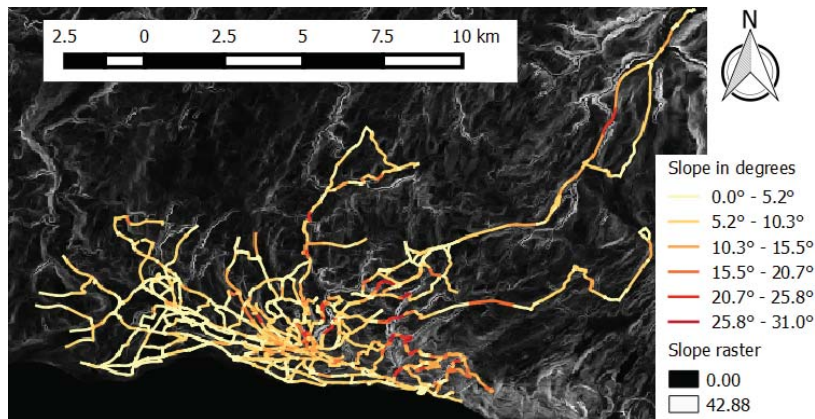
Figure 15.5 – Slope profile of street segments of the city.

Land-use data

As mentioned already, land-use features are measures of types of vegetation, population density, building heights, heating type, terrain elevation, etc. A list of land-use variables relevant to air pollution were selected (see below), using data from the Swiss Federal Office of Topography (Swisstopo) and the Swiss Federal Office of Statistics (OFS). It is worth mentioning that we have extracted these data not only for the street segments that are covered by the mobile sensor network but also for the other parts of the city. This will later enable us to estimate the air pollution for the segments that are beyond our sensor network coverage.

■ **Altitude** and **slope:** These are potentially two important land-use parameters. We obtained a digital height library of Lausanne from Swisstopo that represents the 3D form of the earth's surface with spatial resolution of 25 m and average error of 1.5 m. Applying a digital elevation model (DEM) tool in a GIS software, we extracted the mean altitude and the mean slope of the street segments of the city. Figure 15.4 and Figure 15.5 show the altitude and slope raster data (in the background) and the extracted corresponding
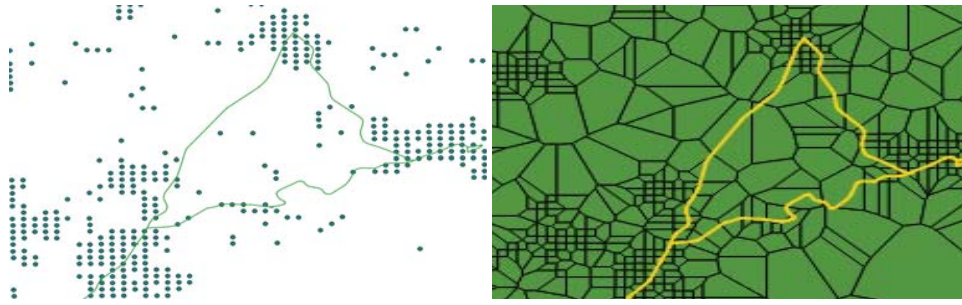
Figure 15.6 – Data points of land-use data in hectare grid format (left) and its corresponding Thiessen polygon partitioning (right). Each point (in the left picture) represents one land-use data point for one hectare. Since the land-use grid data is sparse and non-uniformly distributed, the Thiessen polygon partitioning is used to assign the mean data to the streets segments that traverse a partition.

data of the street segments.

- Density of **population**, **households**, **buildings**, and **industries**: A large variety of land-use and population statistics averaged in 100×100 m^2 (i.e. hectare) grids was obtained from the OFS. Number of inhabitants, buildings, type of fuel used for heating, industry facilities, etc. are the main modalities of this dataset. Table 15.1 (right) gives a complete list of these modalities.

Translating the land-use statistics data (which is in hectare grid format) to our street segments spatial model was not a trivial task, due to the fact that the land-use data is sparse and that not every point in the hectare grid has a value (see Figure 15.6, left). Moreover, street segments are of different lengths, so they traverse different number of grids and each would be corresponding to various land-use statistics values. To produce an adequate predictor based on these data, we first partitioned the whole area of the city using Thiessen polygons. This divided the area in polygons having the value of the closest point with a land-use statistic measurement (shown in Figure 15.6, right). Then, the average of the values of the polygons that were intersected by the street segments was calculated and assigned to each street segment to be used as a predictor.

Traffic data

Traffic has a significant and direct impact on urban air pollution and it is of course desirable, if available, to consider information on its characteristics within an air pollution modeling framework. In our work we used traffic count data provided by Transitec, a company collaborating with the Lausanne Road and Mobility Service.

We extracted five traffic modalities for our street segments, listed in Table 15.2. Figure 15.7 is an example that presents the resulting allocation of one of these modalities (i.e. average daily traffic charge) over the street segments of the city. The first four modalities (in Table 15.2) have been directly provided by the datasets, while the last parameter was produced by us.

Table 15.2 – The traffic modalities extracted for street segments.

Parameter	Symbol
Daily mean charge	TJM
Daily traffic volume mean during working days	TJOM
Hourly charge during morning rush hours	HPM
Hourly charge during night rush hours	HPS
Estimated effect of highways	HT
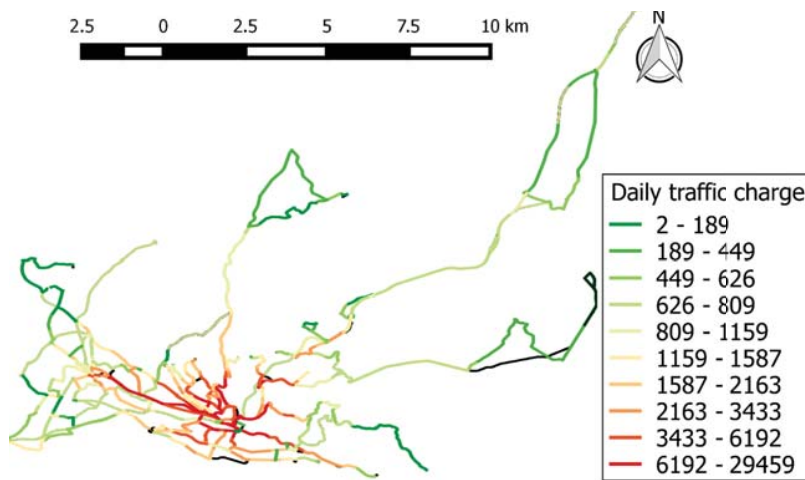
Figure 15.7 – Average daily traffic charge (TJM) of street segments.

Since the traffic volume of the highways are very likely to have an important effect on the LDSA measurements, a variable to model the "estimated effect of highways" was produced as follows:

$$HT_i = \begin{cases} \frac{T_H}{d^2} & \text{if } d \geq d_{min} \\ T_H & \text{if } d < d_{min} \end{cases} \tag{15.1}$$

where T_H denotes the mean traffic charge in the closest highway and d is the distance of the street segment i to that highway. This predictor variable estimates the effect of highways by taking the traffic volume of the closest highway to each street segment divided by the square of the distance to that highway. If the distance is less than d_{min} (a constant thresholding parameter, which was set to $d_{min} = 1$ m), we simply give the value of the traffic volume to this predictor.

Summary

In this chapter, we introduced the problem of generating complete pollution maps from spatio-temporally sparse measurements, and the need for air pollution modeling. We discussed previous work in this field, including both physics-based and statistical modeling. Opting for the latter modeling class in our work, we presented the preliminary considerations of a statistical modeling framework, by discussing our choices in terms of spatio-temporal domain discretization and explanatory variables.

16 Proposed Techniques

W E propose five statistical modeling methods for generating high-resolution air pollution maps in urban scenarios, leveraging measurements from our mobile sensor networks and various explanatory variables. For the first two techniques, we assume proxy pollutant and meteorological explanatory data only, while for the latter two we study the benefit of adding land-use and traffic information to our modeling framework.

16.1 Proxy Pollutant- and Meteorological Data-based Methods

Among the different types of explanatory variables that we considered, pollution time series from traditional monitoring stations and weather data are arguably the most easily available, for many parts of the world. In many cases, these data are made freely available by governmental agencies, which is not typically the case for either land-use or traffic data.

For these reasons, in this section, we discuss statistical modeling methods that make use exclusively of these type of data as explanatory variables. We propose two such modeling techniques: a log-linear regression and a novel network-based log-linear regression.

16.1.1 Log-linear regression

It has been experimentally shown that the mathematical links between gaseous parameters in air are logarithmic [102], [110]. In our first modeling attempt, similar to many previous works (e.g., [67]) we used a log-linear regression model to estimate LDSA values in every street, with the data of proxy pollutants and meteorological explanatory variables, defined in Table 15.1, serving as model inputs. The mathematical formulation of this model is defined by the following equation:

$$\log(L_m) = \alpha_0 + \sum_{i=1}^{9} \alpha_i \cdot \log(v_i) \tag{16.1}$$

where L_m denotes the LDSA estimated value in segment m, α_0 the intercept, v_i the explanatory variable i, and α_i the coefficient of each variable.

We divided the available data into two subsets, the "training set" and the "validation set", using 10-fold cross validation. On the training set, we used the QR decomposition algorithm [111] to solve the linear least squares problem in order to find the coefficients of the model for each street segment. Models were developed for each street on on three time resolutions (daily, weekly and monthly). For the quarterly time resolution, too few aggregated data points were available within the considered interval in order to fit a model per street. More precisely, only four complete quarters are available, which are too few to fit a model with ten parameters. The results of using this modeling approach, which we will henceforth refer to as *Basic Log-Linear (BLL)* regression, are reported in Section 16.3.

16.1.2 Network-based log-linear regression

The goal is to estimate the LDSA values for the location-time pairs that the mobile sensor network has not covered (but still covered previously at least once). The previous model computed the LDSA values based only on the values of the explanatory variables (which are always available through the static stations). Our next idea was to take also into account the measured LDSA values of other segments for predicting the LDSA values in a given segment and in a given time window.

However, this was not trivial considering the fact that the sensors are mobile and the coverage is dynamic, an especially important factor when high temporal resolutions (e.g., hourly) are considered. For instance, if the model of the segment S_m is dependent on the LDSA value of the segment S_n, then the model cannot work when there is no bus covering the segment S_n. To address this challenge we propose to build a *virtual dependency network* on the segments. In this network, each street segment is one node and a directed edge is drown between node S_m and S_n if node S_n is considered as a variable in the model of node S_m. As we will see this network is able to address the problem of dynamic coverage of the mobile sensors.

Now the question is how to build the network and define how the models work on the network. We propose to connect node S_i to S_j if the following two conditions hold:

1. The Pearson correlation between the LDSA values of node S_i and S_j is high. This is due to the fact that correlation is a basic need for every variable in a model. In Chapter 13, we have already discussed that the segments which are geographically close do not necessary show high correlations, especially if they are not in the same direction (see Figure 13.2).
2. Node S_j has reported enough *complementary data* relative to the available data of node S_i. We define complementary data as the number of time-slots when there are LDSA values reported for the segment S_j but no value was reported for S_i. Based on the first condition, some of the segments which are along one street have the highest

Figure 16.1 – The generated network for Lausanne's street segments for LDSA estimation model. Each red node represents the center of one segment while the black lines between two nodes represent the correlation between the LDSA values of two segments. The big 9 blue nodes represent the nine explanatory variables which are in fact connected to all other nodes. For increasing the readability of this figure, we have removed all the edges between the blue nodes and the red nodes.

correlations and since they are most probably covered by the same bus they do not have any complementary data, meaning that either all segments have data or none of them (making the models inefficient). With this second condition we make sure that the selected network edges are going to be useful for the models.

To create the network based on the two mentioned edge conditions, using the available LDSA data of the segments, we computed the cross correlation of LDSA values of all combinations of the segments and then for every segment S_i we find the M most correlated segments. M was set to 10 for all results reported in this work. Among the most correlated segments we find the ones that have more complementary data relative to S_i. We establish an edge for every node S_i and then keep adding edges to the network (while considering the two conditions) until the network is connected.

To take the nine explanatory variables into account, we insert them as nodes to this network and connect them to every other node in the network. In our deployment, this process generated a network with 1386 nodes (1377 street segments + 9 explanatory variables) and found 15 040 edges. Figure 16.1 shows this network.

Since we have ten buses, during one time slice (e.g., one hour) a fraction of nodes of the network will have actual measurements, while the rest should be estimated using the models. It is obvious that if the network is (at least) minimally connected, then theoretically all the nodes can be predicted one by one even if only one segment has actual measurement. Denoting E as

the edge list of this network, and L_{S_m} as the LDSA value of segment S_m, here is the proposed model for node S_m:

$$\log(L_{S_m}) = \alpha_0 + \sum_{i=1}^{9} \alpha_i \cdot \log(v_i) + \sum_{[m-n] \in E} \alpha_n \cdot \log(L_{S_n}) \tag{16.2}$$

The optimal values for the coefficients of this model are found similarly to the first method using the training set data.

This novel graph-based approach, which we will refer to as a *Network-based Log-Linear (NLL)* regression, iteratively estimates the LDSA value of the segments based on the LDSA value of the correlated nodes and also based on the values of the explanatory (meteorological and gaseous) variables. This is an approach well-suited for mobile sensor networks where the coverage area of the network dynamically changes over time.

Theoretically, it does not matter which nodes have measurements and which nodes are to be estimated, as long as there is at least one node with a measurement, the LDSA value of all the other segments can be estimated iteratively. However, the estimation accuracy in the case of multiple iterations will drop. The results presented in Section 16.3 were obtained by running the NLL algorithm for only one iteration.

16.2   Methods that Exploit Land-use and Traffic Data

Both of the previous methods have an important drawback: they can only be used to estimate pollution within a certain segment if it is within the coverage area of the mobile sensor network (i.e. it is part of the public transportation network). However extensive, a public transportation network does not cover every single street within a given city.

Although more difficult to attain and to process, land-use data, in particular, but also traffic data (if complete), can help overcome this issue, by extending the estimating power of statistical models to urban areas that, although not covered by the mobile sensor network directly, have a high degree of similarity in terms of land-use and traffic characteristics with the ones that are.

We present here three modeling techniques that use an extended set of explanatory variables, including the land-use and traffic parameters presented in Chapter 15, as well as meteorological and proxy pollutant data.

16.2.1   Log-linear regression with land-use

The first model we propose, which includes land-use and traffic data, is a direct extension of the basic log-linear regression model presented in Section 16.1. We label it a *Basic Log-Linear*

with Land-Use (BLL-LU) regression. It is described mathematically by the following equation:

$$\log(L_{S_m}) = \alpha + \sum_{i=1}^{9} \beta_i \cdot \log(v_i) + \sum_{i=1}^{17} \gamma_i \cdot \log(U_{i,S_m}) \tag{16.3}$$

where U_{i,S_m} denotes the land-use and traffic variable i in segment S_m (see Table 15.1 and Table 15.2 for the list of variables), α the intercept, and β and γ the coefficients of each variable. Hereafter, the operator 'log' actually represents a 'shifted-log' transformation due to many zeros in environmental data (a positive offset, in this case a value of one, has been added to the data before log-transforming, e.g, $\log(1 + v_i)$).

Since the land-use and the traffic data inherently provide spatial variety, we are able to build one general model for the city, which can provide estimations for regions outside our sensor network coverage, as long as land-use and traffic information is available.

16.2.2 Land-use network-based log-linear regression

Differently from the network-based modeling method presented in the previous section, we now propose building the virtual dependency network using land-use data, and also adding land-use and traffic data as explanatory variables. We call this a *Land-Use Network-based Log-Linear (LU-NLL)* regression.

The mathematical formulation of the model for this technique is an extension of Equation (16.3):

$$\log(L_{S_m}) = \alpha + \sum_{i=1}^{9} \beta_i \cdot \log(v_i) + \sum_{i=1}^{17} \gamma_i \cdot \log(U_{i,S_m}) + \sum_{[m-n] \in E} \delta_n \cdot \log(L_{S_n}) \tag{16.4}$$

where the last term in Equation 16.4 indicates that we have used the LDSA of street segment n as an input to estimate the LDSA of street segment m. δ denotes the coefficient for this LDSA input.

At this point, the question is how to define the dependency network. Instead of defining the dependency network based on the LDSA correlations of the street segments, we define it based on the land-use data. This provides a significant advantage since the dependency graph (based on land-use data) no longer requires LDSA measurements and can be extended to all the streets of the city (beyond the coverage of the mobile sensor network). The land-use data are used in this model as particular distinguishing features of the street segments and naturally categorize them into similar classes.

In this graph, each street segment is one node and we connect the nodes with highest similarity in terms of their land-use data. Therefore the criteria for connectivity of nodes
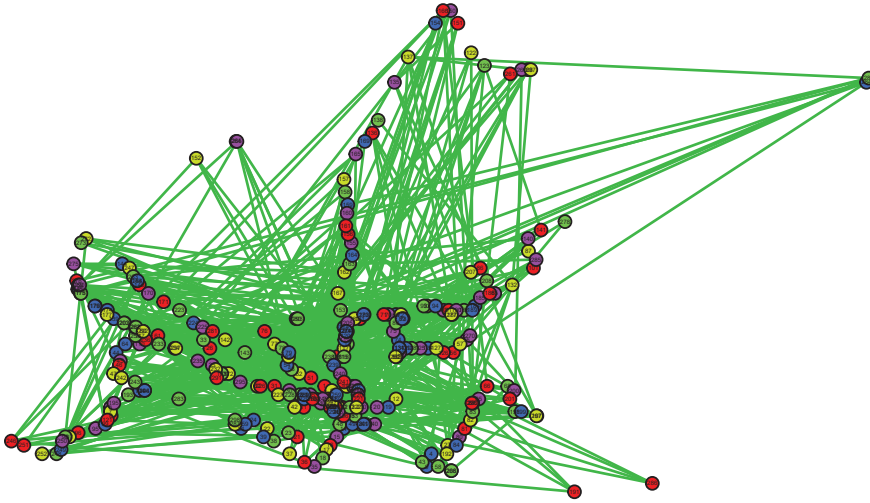
Figure 16.2 – Virtual land-use graph generated for the street segments of the city. Each node represents one street segment and the edges are drawn based on similarity in the land-use data. For increasing readability, not all the nodes and edges are shown in this figure.

in the dependency graph is defined as follows:

$$D_L(S_m, S_n) = \sqrt{\sum_{i=1}^{17} \left(\widehat{\log}(U_{i,S_m}) - \widehat{\log}(U_{i,S_n}) \right)^2} \tag{16.5}$$

where $\widehat{\log}$ denotes a normalized shifted-log transformation. $D_L(S_n, S_m)$ is the Euclidian land-use distance between two arbitrary street segments S_m and S_n. We calculate this distance for all combinations of street segments and then we connect each street to the one with the smallest distance. We keep connecting nodes with smallest land-use distance until all the graph becomes connected. Figure 16.2 shows partially such a graph generated for the city under study.

To the best of our knowledge no previous work has ever built a virtual network based on land-use data and integrated it into such a model.

To train the model, we again divided the available data into a training set and a validation set, using 10-fold cross validation, and solved the the linear least squares problem in order to find the coefficients of the model. Working on four time resolutions we developed four models for the whole city (differently from Section 16.1 in which one model per street was generated).

16.2.3   Deep learning model

For all the previously proposed regression models we used an assumption of (log-)linearity between the LDSA and the model input variables. In fact, if the correlations between the input variables and the target variable (i.e. LDSA) are (log-)linear, then the regression model

can estimate the target pollutant very well. However, this might not always be the case and and in this subsection we investigate a more flexible approach, that allows the model to capture potential non-linear and complex relationships between the variables. To this end, we consider the opportunity of using an artificial neural network (ANN) model. We refer to this approach as a *Deep Learning Model (DLM)*.

ANNs are usually useful for automatically capturing nonlinear dependencies between input variables, without needing to know the underlying function. Theoretically, an ANN function is defined as a composition of other functions which can further be defined as a composition of other functions. This can be conveniently represented as a network structure, with arrows depicting the dependencies between functions and variables. The universal approximation theorem [112] states that a feed-forward network with a single hidden layer containing a finite number of neurons (also called a Multi-Layer Perceptron, MLP) can approximate any continuous function under mild assumptions.

Studying the history of multi-layer neural networks, their difficult optimization has mostly prevented obtaining expected benefits of going beyond one or two hidden layers [113]. However this situation has recently changed with the promising approach of Deep Neural Networks (DNNs) and stacked AutoEncoders (stacked AEs) [113]–[116]. A key component of this success is the procedure of partially pre-training the network using an unsupervised training criterion.

The input data itself is used to partially train the neural network. This way, the network can learn intrinsic information about the data without the help of a target vector (i.e., the measured LDSA data in our case). In this process, the network automatically learns high-level features from the input data and it captures the nonlinear dependencies between them. The learned information is stored as the initial weights of the network.

Moreover, through this unsupervised learning process the dimensions of the input data is reduced while the amount of information in the data is not significantly lost. AEs and restricted Boltzmann machines are two equally efficient methods to generate and pre-train DNNs. In our work, we have used AEs to automatically capture the useful features of the data.

Automatic feature extraction using AEs

AEs can be considered as multi-layer sparse coding networks. The differences between AEs and MLPs are that (i) in an AE the output layer has the same number of nodes as the input layer, and (ii) instead of being trained to predict the target value given a set of inputs, AEs are trained to reconstruct their own inputs.

In other words, AEs try to learn an approximation to the identity function, such that the output $\hat{x}$ is similar to the input x. By placing constraints on the network, such as by limiting the number of hidden units (z), AEs can discover interesting structures about the data.
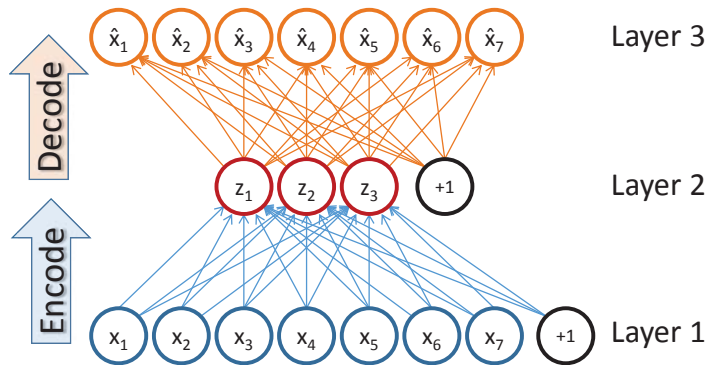
Figure 16.3 – An exemplar autoencoder with three layers. The outputs ($\hat{x}_i$) are learned to predict the input variables (x_i). After training, z (layer 2) will be an encoded representation of x with lower dimension.

An AE first maps the scaled d-dimension input vector $x \in [-1,1]^d$ to a hidden d'-dimension representation $z \in [-1,1]^{d'}$ (where $d' < d$) through a deterministic mapping function:

$$z = f_\theta(x) = s(Wx + b), \tag{16.6}$$

parameterized by $\theta = \{W, b\}$, where s is the activation function (e.g., sigmoid, tanh, softmax), W is a $d' \times d$ weight matrix, and b is a bias vector. The resulting latent representation z is then mapped back to a "reconstructed" vector $y \in [-1,1]^d$ of the input vector, thus:

$$\hat{x} = g_{\theta'}(z) = s(W'z + b') \tag{16.7}$$

with $\theta' = \{W', b'\}$. The parameters of this model (θ and θ') are optimized to minimize the average reconstruction error:

$$(\theta^*, \theta'^*) = \underset{\theta,\theta'}{\arg\min} \frac{1}{d} \sum_{i=1}^{d} (x_i - \hat{x}_i)^2 = \underset{\theta,\theta'}{\arg\min} \frac{1}{d} \sum_{i=1}^{d} \left(x_i - g_{\theta'}\left(f_\theta(x_i)\right) \right)^2 \tag{16.8}$$

Eq. (16.8) shows that θ^* and θ'^* are functions of only the input vector x, implying that a completely unsupervised back-propagation learning approach can learn optimized weights.

Figure 16.3 shows an example of an AE network with three layers. In this example, there are seven inputs x_i and there are 3 hidden units (z_i) in the middle layer, so the AE network is forced to learn a compressed (low-dimensional) intermediate representation of the input and then re-construct the 7 input data ($\hat{x}_i$) using z in an unsupervised manner. If there are structures in the data, for example, if some of the input features are correlated, then the AE will be able to discover some of those correlations.

Once the weights of the AE network are learned (through an optimization method, in our work AdaGrad explained later in this subsection), the "decode" part of the AE network (colored with orange in Figure 16.3) is eliminated and the middle layer (z) is set as the input of the next layers
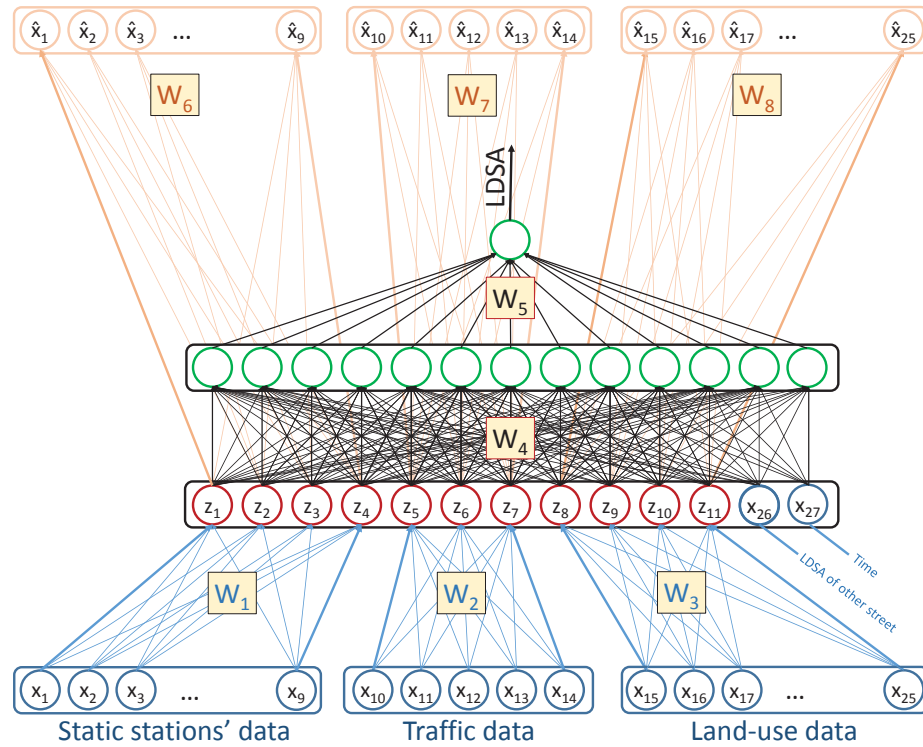
Figure 16.4 – The proposed DNN structure. The blue nodes ($x_1 - x_{28}$) represent the input data, the red nodes ($z_1 - z_{11}$) represent the lower dimension features of the input data and the orange ($\hat{x}_1 - \hat{x}_{25}$) shows the decode part of the AEs.

of a feed forward DNN which tries to model the target modality. The obtained coefficients of the AE are used as the initial values of the weights of the larger multi-layer DNN.

The structure of the deep network

Our proposed network structure uses AEs for each of the different sources of explanatory data to (i) extract informative features of the data, and to (ii) lower the dimensions of the input data. Figure 16.4 presents the proposed structure of the DNN. For increasing the readability we have not shown the usual 'bias' nodes in each layer. The details of this structure are as follows:

1. The nine modalities of the meteorological and air-quality measures obtained from the two static stations (listed in Table 15.1, left) are coded into a feature vector with four nodes ($z_1 - z_4$).

2. The five measures related to traffic (listed in Table 15.2) are mapped into a feature vector with three nodes ($z_5 - z_7$).

3. The eleven measures related to land-use (listed in Table 15.1, right) are mapped into a feature vector with four nodes ($z_8 - z_{11}$).

135

4. In order to take into account the land-use network shown in Figure 16.2, we have inserted node x_{27} into this DNN. This node represents the LDSA measurement of the most similar street segment in terms of land-use data. Since this is only one node, there is no need to propose an AE for it.

5. The last modality taken into account is "time", represented by x_{28} in the DNN. The value of this node is dependent on the time resolution of the model. Since four time resolutions (daily, weekly, monthly, and quarterly) have been considered, this nodes has four different implication (i.e., day number, week number, month number, or quarter number).

In each intermediate layer of the DNN, we used the following *nonlinear* neural transfer function to compute the output of the nodes:

$$h_{W,b}(x) = \tanh(\sum_{i=1}^{N} W_i x_i + b) \tag{16.9}$$

where x_i represents the input nodes, N is the number of inputs, while b and W_i are the bias and the weights to be optimized. The hyperbolic tangent that is used as activation function is suitable for our model due to the fact that all the inputs and the target modalities are scalar numbers that are normalized in $[-1, 1]$.

Training and implementation

We used the same exact training set, validation set and land-use network explained in subsection 16.2.2 also for this model. Using the training set, first the AEs of the DNN were trained (unsupervised), and optimized values of their weights ($W_1, W_2, W_3, W_6, W_7, W_8$) were obtained. Then the upper part of the AEs (colored orange) was discarded and the whole DNN was trained (supervised) starting with the partial initial weights' values obtained from the AEs.

For optimizing the weights and the biases, we used a recent adaptive sub-gradient optimization method called AdaGrad [117] which has shown remarkably good results on large scale multi-dimensional learning tasks in a distributed environment [118]. AdaGrad automatically scales the learning rate of each parameter in every iteration and tries to dampen the extreme parameter updates, while increasing the learning rate of parameters that have got few or small updates. The loss function used as the criteria of the optimization process was simply the root square error of the estimated values relative to the measured values.

We implemented our models in the framework of Tensorflow [119] which is an open-source framework for numerical computations using data flow graphs, recently released by Google. This framework significantly facilitates the process of building and training the networks and provides useful optimization libraries.
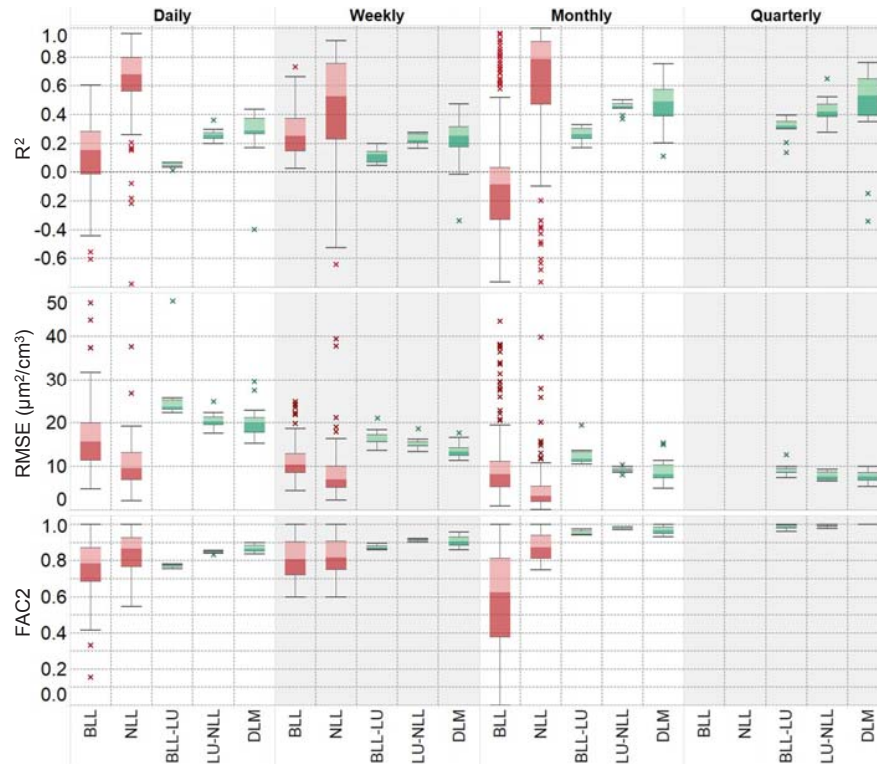
Figure 16.5 – The evaluation results for the proposed modeling methods. The results of the methods that derive one model per street (BLL and NLL) are shown with red boxplots, while the results for the other modeling techniques are depicted with green boxplots.

16.3 Performance

All five modeling methods we have introduced in this chapter have been tested using the same data set, with models being derived for daily, weekly, monthly, and, in the case of models that exploit land-use and traffic data, also quarterly time resolutions. Denoting M as the set of modeled values and O as the set of corresponding observations, we consider the following three metrics for performance evaluation:

1. The **coefficient of determination** (R^2) shows the proportion of the variance in the model output that is predictable from the explanatory variables. This metric is defined as:

$$R^2 = 1 - \frac{\sum_i (O_i - M_i)^2}{\sum_i (O_i - mean(O))^2} \tag{16.10}$$

   where $mean(O)$ represents the mean of all observations which are considered in the validation sets. $R^2 = 1$ indicates that the fitted model explains all variability in the model output. A negative R^2 indicates that the fitted model is even worse than a simple mean of the data.

2. The **Root Mean Square Error** (RMSE) is computed as the following:

$$RMSE = \sqrt{\frac{1}{L} \sum_i (O_i - M_i)^2} \tag{16.11}$$

where L is the number of estimations provided by the model. Obviously, the lower the RMSE is, the better the model works.

3. The **factor of two measure** (FAC2) is the percentage of ratios O_i / M_i that fall between 0.5 and 2:

$$0.5 < \frac{O_i}{M_i} < 2 \tag{16.12}$$

The closer the value of this metric is to 1, the better the model has estimated the observations.

The results of calculating these metrics for all the proposed models are shown in Figure 16.5. Although not directly comparable, some observations can be made on the difference between the attainable performance for methods that derive models for each street (i.e. models that do not use spatial explanatory variables), and models that make use of land-use and traffic data to derive city-wide models. As can be expected, the BLL and NLL methods are able to deliver locally higher performance than a single model for the whole network. The variability of this performance is, however, large. Also, as the temporal aggregation is increased, these methods start having a harder time fitting their local models because of the reduced number of aggregated data points within the 14 months considered. This starts affecting their performance at the monthly scale, while making it impossible to fit local models at the quarterly levels (there are only four complete year-quarters in the considered interval).

The three models that also exploit land-use and traffic data (BLL-LU, LU-NLL and DLM) do not suffer from this problem, as a single model for the whole city is derived using data from the all the segments in the covered street network. While their performance is moderate (e.g., R^2 values higher than 0.5 are only achieved for the lower temporal resolutions), these techniques have the significant advantage of providing a single model for the whole city which enables the generation of high-resolution maps for all regions where land-use and traffic data is available.

Figure 16.6 shows two examples of daily air pollution maps of the city generated by the proposed DLM technique. They include streets that are outside the direct coverage of our mobile sensor network.

For both cases when a virtual dependency network is used (i.e. for the NLL and LU-NLL methods) this has a significant positive impact on performance when compared to the basic log-linear regression methods. Finally, the DLM technique is able to surpass the LU-NLL, delivering the best performance out of the single-model class of techniques, for all the considered metrics.
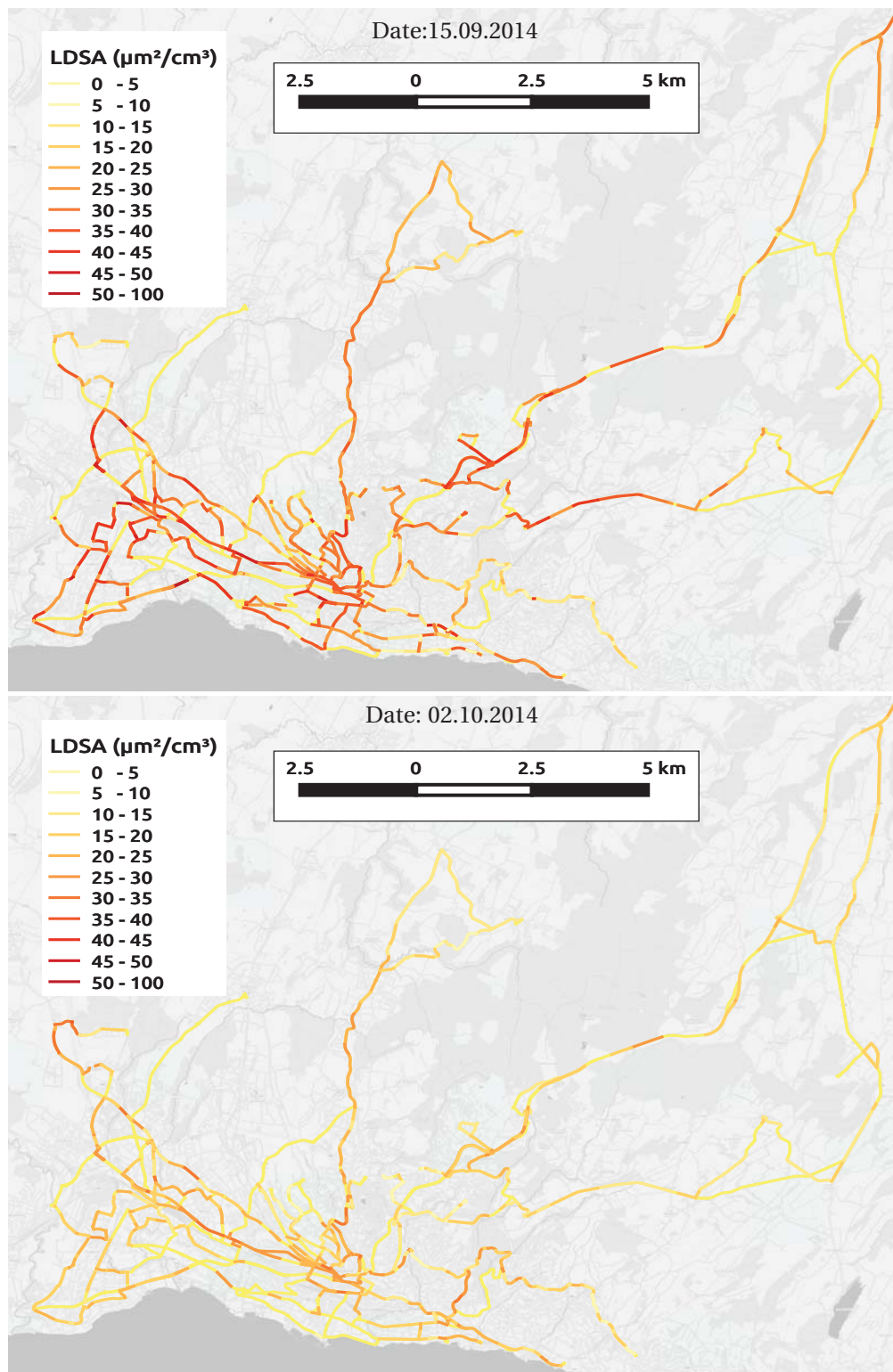
Figure 16.6 – Two examples of daily LDSA maps generated by the DLM approach.

Summary

In this chapter, we have presented five different statistical methods for generating high-resolution air pollution maps using measurement data from our mobile sensor network. The first two techniques make use of proxy pollutant and meteorological information as explanatory variables, in order to derive local air pollution models for each segment covered by the sensor network. The other three techniques additionally exploit land-use and traffic data to derive a single air pollution model for the whole city. The main benefit of these latter techniques is that they allow us to build complete maps, going beyond the coverage of the mobile sensor network. We evaluated and reported the performance of each of these techniques for multiple temporal resolutions, using three different statistical metrics.

17 Conclusion

H IGH-RESOLUTION urban air pollution maps are needed in order to push our understanding of the health impact of pollutant exposure at an individual level. Mobile sensor networks hold the potential to enable such detailed urban maps by increasing the spatial resolution of measurements. However, due to the inherently dynamic and partial nature of the coverage attainable with such a system, a modeling framework is needed to be able do generate complete pollution maps.

In our work, we investigated the use of statistical modeling techniques to generate high-resolution air pollution maps. Using more than 44 millions of geo- and time- stamped LDSA measurements (i.e. 14 months of real data), we proposed and studied two different types of modeling techniques: (i) methods that use only proxy pollutant and meteorological explanatory variables from static stations to generate local pollution models, and (ii) methods that also integrate spatial explanatory variables (i.e. land-use and traffic data) to derive unified city-wide models. In this chapter, we conduct a critical discussion of the performance of these techniques, placing our contribution in a broader context, and outlining future directions for improvement.

17.1   Proxy Pollutant- and Meteorological Data-based Methods

We proposed two log-linear regression methods (BLL and NLL) which, apart from our mobile sensor network data, made use of only proxy pollutant and meteorological data from the NABEL and, respectively, MeteoSwiss monitoring stations. While the use of log-linear regression models for building high-resolution pollution maps based on mobile sensor network data is not new [67], we are the first to integrate meteorological time series explicitly, in a mobile sensing context. Furthermore, our virtual dependency network approach is completely novel, and the benefit of using it in order to add mobile measurements from correlated and data-complementary street segments as model inputs, is apparent from the presented results.

Using both modeling methods, we were able to generate pollution models for the street

segments covered by our mobile sensor network for daily, weekly and monthly time resolutions. Due to the high number of model parameters compared to the available data samples, quarterly models could not be built for the considered time interval. Data for multiple deployment years are required in order to be able to derive models at this time scale.

As both the BLL and NLL regression techniques work by deriving a model for each considered street segment, a high fitting performance can be achieved locally. However, this performance is characterized by a relatively high variability. The source of this variability needs to be further studied. Possible causes include the non-uniform density of measurements per street segment, but also a potentially varying degree of correlation between the considered static monitoring station explanatory variables and the considered street segment.

Finally, the biggest weakness of both of these approaches is their inability to estimate pollution levels for street segments outside of the mobile sensor network coverage. This limitation cannot be overcome without the use of spatial explanatory variables available beyond the coverage network such as land-use or traffic data.

17.2 Methods that Exploit Land-use and Traffic Data

We proposed three different modeling methods (BLL-LU, LU-NLL and DLM) able to leverage land-use and traffic count data for generating high-resolution urban air pollution maps. The idea of using land-use data for air pollution modeling in the context of mobile air pollution monitoring networks is not novel, having been addressed in the works of both Hasenfratz et al. [67] and Li et al. [77]. Neither of these works, however, used temporal explanatory variables, which implied deriving different pollution models for each targeted period (e.g., one spatial model for one specific day).

By using both temporal and spatial explanatory variables we are able to build a single unified city-wide model for each considered temporal scale. In other words, the daily model obtained through any of our three techniques can be used to estimate pollution for any day for which explanatory data is available. The resolution and the area coverage of the maps generated by the proposed models are beyond the ones of our mobile sensor network, and are extended to the resolution and coverage of the land-use and traffic data.

As we have seen also with the previous class of models, the use of a virtual dependency network has a significant impact on improving model performance compared to the BLL-LU technique (for all considered metrics). Differently from NLL, the LU-NLL approach builds the dependency network based directly on the land-use data. This significantly simplifies the problem of generating a connected graph, as no condition on the complementarity of the mobile measurement data is required.

The DLM method, based on Deep Neural Networks, outperforms both BLL-LU and LU-NLL, implying that its added complexity is adding significant value in terms of estimation

performance. Both the LU-NLL and DPL are completely novel approaches to generating high-resolution air pollution maps based on mobile sensor network data.

Although a generally good, and very consistent, performance was obtained for all of the three methods that used spatial explanatory variables, the local performance obtained by both BLL and NLL, at daily and weekly time scales, is higher. As such, an option for maximizing the quality of the maps could be to consider a hybrid approach: using a local modeling approach wherever this is available, and the fitting quality is high, and using a global model where this is not the case (including the street segments not covered by the mobile sensor network).

Another potential direction for improving the performance of modeling techniques that exploit land-use data, lies in the refinement of the way predictors are mapped to street segments. This, as we have seen in Chapter 15, is not a trivial task.

In the future, we plan to apply similar modeling methods for generating high-resolution maps for the other measured modalities (e.g., CO, NO_2, and O_3). However, as previously mentioned, the problems of mitigating the distorting effect of mobility and calibration need to be appropriately addressed for each of them before proceeding to the modeling stage.

Summary

We conclude our work on high-resolution air pollution mapping with a critical discussion of the performance attained with the proposed modeling techniques. We place our contributions in context of related work and discuss future directions for improvement. We have proposed five statistical modeling techniques, three out of which (NLL, LU-NLL and DLM) are original contributions. We discussed the advantages and drawbacks of local and global modeling techniques and proposed the option of using a combination of the two (i.e. a hybrid technique), that would allow us to jointly exploit their benefits.

Conclusion Part VI

18 Conclusion

T HIS dissertation investigates the potential of mobile sensor networks to enhance urban air quality monitoring by pushing the spatial resolution attainable with traditional measurement systems. This endeavor was motivated by a need to better understand the health impact of air pollution on urban populations and recent technological advances in the field of affordable miniaturized air quality sensors. The end-objective of a mobile air quality monitoring network is to deliver high-resolution air quality maps which can then be used by health specialists, officials, or, even, private citizens. However, achieving this task presents many scientific and technical challenges. These include questions on: (i) system design, deployment, and operation, (ii) mitigating the distorting effect of mobility on low-cost chemical sensor measurements, (iii) achieving and maintaining sensor calibration over the lifetime of the deployment, and (iv) generating complete air pollution maps, considering the dynamical coverage of the mobile sensor network. In this manuscript, we presented our work on each of these topics. Through it, we hope to have provided enough evidence that mobile sensor networks are a valid and promising approach for delivering a deeper understanding of urban air pollution.

18.1 Summary of the Work

Our work on designing mobile air quality monitoring networks for urban environments provided the following contributions:

- We developed a successful mobile sensor network for monitoring air quality, which has been functional for more than 33 months to date.

- We presented a step-by-step guide to its design in this manuscript. This included the choice of targeted pollutants, sensor selection, node design, server architecture, and deployment management.

- We proposed a novel modular node architecture which provides flexibility for investigating future system upgrades (e.g., new sensor modalities, different air sampling

systems).

Our work on mitigating mobility-caused measurement distortions on low-cost chemical sensors investigated both signal processing and mechatronic techniques. Our contributions on this topic are summarized as follows:

■ We presented a rigorous wind tunnel-based experimental framework for studying the distorting effect of mobility on commercially available chemical sensors, that typically suffer from slow dynamic response (i.e. long response times).

■ We showed that, even at low platform movement speeds, this effect is not negligible.

■ We proposed the use of signal deconvolution, using a Wiener filter and a dynamic sensor model, for reconstructing the underlying true signal. Using the same experimental framework we showed the effectiveness of this method.

■ Since the effectiveness of the deconvolution technique is limited by the signal-to-noise ratio of the raw measured signal, we investigated the possibility to enhance it through a mechatronic approach (i.e. the use of an active air sampling system). We showed that an appropriately designed active sniffer can indeed improve the signal-to-noise ratio of the measured signal, through both wind tunnel and outdoor experiments.

Our work on the calibration of mobile low-cost chemical sensors focused on single-hop rendezvous calibration (i.e. between a static reference station and a mobile sensor node). Our contributions are summarized as follows:

■ We investigated the opportunity of using more complex sensor models, which are more faithful to their complex behavior. As such, we proposed a model-based approach to rendezvous calibration and showed the potential for improving calibration performance when using complex models over larger calibration time windows.

■ Building upon our work on mitigating the distorting effect of mobility on chemical sensors, we proposed a mobility-aware rendezvous calibration technique which integrates a pre-processing step in which the mobile measurement data is deconvolved. Using real mobile measurement data from our deployment we showed the potential of this technique to improve calibration accuracy.

Finally, on the topic of building high-resolution air pollution maps based on our mobile sensor network measurements, we focused on the particulate matter modality and proposed the use of statistical modeling for addressing the dynamic coverage of the system. We divided the spatial domain into street segments, and considered models at different time resolutions (e.g., daily, weekly, monthly, quarterly). Our contributions are summarized as follows:

■ We proposed two techniques that use proxy pollutant and meteorological time series as explanatory variables. The first one is a basic log-linear regression using solely these inputs and the particulate matter data as input. The second one is a novel network-based log-linear regression technique, which uses a virtual dependency network to add,

as model input, particulate matter measurements from street segments that have a high Person correlation and a high degree of data complementarity with the considered street segment. Using both methods, we were able to build local (i.e. per street) pollution models for the streets that were within the coverage of our mobile sensor network. Analyzing their performance, we showed the benefit of using a virtual dependency network in the second technique.

- Seeking to expand our air pollution maps also to regions that are not directly covered by our mobile network we added land-use and traffic count data to the set of explanatory variables. We, thus, extended both of the previous regression techniques to include these new inputs. We also modified the way the virtual dependency network was built, by considering connections between street segments that have similar land-use characteristics. Using both methods we were able to derive unified city-wide models for each of the considered time scales. In this case also, the performance of the network-based approach was superior to the basic log-linear regression.

- Using all of the available explanatory variables and the virtual dependency network, we proposed another novel technique for deriving complete urban pollution maps by employing artificial neural networks with a deep learning approach. This method provided the highest performance of all the proposed modeling techniques that derived city-wide models.

18.2   Discussion and Outlook

Considering the novelty of the domain, this work represents some of the first steps towards enabling high-resolution air quality monitoring through the use of mobile sensor networks. This is true for all the different topics we considered in this manuscript, each one of which needing to be appropriately addressed in order to ensure the long-term success of this type of application.

A poor system design can jeopardize deployment reliability. Our sensor network deployment in Lausanne has been operational for over two years, and although some adaptations of various subsystems needed to be made along the way, the network as a whole was never shutdown. Nevertheless, the opportunity of adding other sensor modalities, such as local wind measurements needs to be further investigated. Local wind patterns play an important role on pollutant dispersion, however, the choice of an appropriate low-cost technology to be incorporated into a highly mobile sensor node can be challenging in itself. On the network level, suitable automatic diagnosis techniques need to be researched and implemented in order to increase system maintainability, reduce running costs and ensure scalability.

An inadequate platform design can also affect measurement quality and exacerbate mobility effects on chemical sensing. For this reason, we have put an emphasis on enhancing measurement quality through system design. Nonetheless, although we considered multiple sniffer configurations and types of actuation, their design was not optimized. A higher improvement

of the measurement quality can be expected if a more systematic approach to the design process is undertaken. When considering deconvolution for signal reconstruction, a relaxation of the assumptions we have made on the dynamic sensor model needs to be studied. As sensor rise and recovery times have been shown to be asymmetrical, the use of a two phase sensor model might be more appropriate. However, this will certainly complicate the implementation of a deconvolution method.

Online automatic techniques are the most practical way of ensuring calibration over the lifetime of a deployment, and we have made an effort on further advancing this research front. However, both our model-based and our mobility-aware rendezvous approaches still need to be tested in multi-hop calibration scenarios. Furthermore, the opportunity of relaxing the strict rendezvous definition we considered needs to be investigated. In the case of the mobility-aware approach, a practical solution for estimating the dynamic model parameters also needs to be studied.

At this point, we want to clearly highlight the fact that, without suitable solutions for both mobility effect mitigation and online calibration, achieving accurate air pollution mapping for gas-phase pollutants will not be possible. In our work on high-resolution pollution mapping, we have so far targeted solely the particulate matter modality. Nevertheless, we believe that we have now assembled all the necessary elements such that the next natural step would be to expand towards gaseous modalities. This comment is valid also more generally, since, for practical reasons, for each of the topics covered in this thesis, we focused on specific sensor modalities. However, all the proposed techniques need to be evaluated and validated in a broader context, considering diverse sensing technologies.

Looking closer into the topic of high-resolution pollution mapping, the performance of the techniques we proposed can probably be further improved by a refinement of the way explanatory variables are introduced to the considered models. While for meteorological, proxy pollutant, and even traffic count data, the question of integrating them into a statistical model is rather straightforward, this is not the case for land-use data. The mapping between the sparse and differently discretized land use data to street segments is not trivial and should be further studied. Also, most of the modeling techniques we presented are still relatively basic, and the use of more complex learning and probabilistic modeling techniques should be investigated.

While taking into consideration all of the above potential directions for improvement, we do believe that our work has put into place the fundamental pieces of the puzzle necessary to ensure the success of this type of distributed monitoring application, creating a baseline for future developments in this field. In fact, we expect to see a very dynamic evolution of mobile sensor networks for air quality monitoring over the next decade, moving from the current state of academic research towards permanent real-world systems that will provide novel and rich services.

The most basic form of service would provide high spatial resolution air quality maps, which

will become as easy to access on different platforms as weather forecasts are today. The level of complexity of these maps will need to be tailored to the specific customer class targeted by the applications, emphasizing accurate and quantitative information for scientific or regulatory institutions, while placing more focus on intuitive and qualitative information for the general public.

Building upon urban air quality maps, a layer of secondary services is easily imaginable. It would target topics such as: providing advice on outdoor activities (e.g., jogging or biking paths), enabling a new dimension for informed decisions on property rentals or purchases, and, in collaboration with medical specialists, enabling personalized health recommendation systems.

On the platforms side, we expect that, at least for the foreseeable future, public transportation anchored devices will play a central role in ensuring a baseline level of service quality, providing large spatial coverage and a calibration backbone. However, as the sensor technologies advance and costs decrease, we expect that affordable portable platforms will play an increasingly larger role, democratizing the data acquisition process and pushing further the achievable resolution of air pollution maps through crowdsourcing.

Bibliography

[1] World Health Organization (WHO). (2014). 7 million premature deaths annually linked to air pollution - News release. Accessed: 27 March 2016, [Online]. Available: http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/.

[2] R. Berkowicz, O. Hertel, S. E. Larsen, N. N. Sørensen, and M. Nielsen, *Modelling traffic pollution in streets*. Ministry of Environment and Energy, National Environmental Research Institute, 1997.

[3] Enivronmental Protection Agency (EPA). (2016). Criteria Air Pollutants. Accessed: 27 March 2016, [Online]. Available: https://www.epa.gov/criteria-air-pollutants.

[4] European Enivronment Agency (EEA). (1998). Assessment and Management of Urban Air Quality in Europe. Accessed: 27 March 2016, [Online]. Available: http://www.eea.europa.eu/publications/environmental_monograph_2006_5.

[5] European Enivronment Agency (EEA). (2014). Air quality in Europe - 2014 report. Accessed: 28 March 2016, [Online]. Available: http://www.eea.europa.eu/publications/air-quality-in-europe-2014.

[6] Enivronmental Protection Agency (EPA). (2010). Integrated Science Assessment for Carbon Monoxide. Accessed: 28 March 2016, [Online]. Available: https://www.epa.gov/isa/integrated-science-assessment-isa-carbon-monoxide.

[7] Enivronmental Protection Agency (EPA). (2014). Air Sensor Guidebook. Accessed: 28 March 2016, [Online]. Available: https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=277996.

[8] M. Riediker, R. Williams, R. Devlin, T. Griggs, and P. Bromberg, "Exposure to Particulate Matter, Volatile Organic Compounds, and Other Air Pollutants Inside Patrol Cars", *Environmental Science & Technology*, vol. 37, no. 10, pp. 2084–2093, 2003.

[9] Enivronmental Protection Agency (EPA). (2008). Integrated Science Assessment for Oxides of Nitrogen – Health Criteria. Accessed: 29 March 2016, [Online]. Available: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=194645.

[10] S. Grice, J. Stedman, A. Kent, M. Hobson, J. Norris, J. Abbott, and S. Cooke, "Recent trends and projections of primary NO 2 emissions in Europe", *Atmospheric Environment*, vol. 43, no. 13, pp. 2154–2167, 2009.

Bibliography

[11]   Enivronmental Protection Agency (EPA). (2008). Integrated Science Assessment for Sulfur Oxides – Health Criteria. Accessed: 29 March 2016, [Online]. Available: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=194645.

[12]   Enivronmental Protection Agency (EPA). (2013). Integrated Science Assessment for Ozone and Related Photochemical Oxidants. Accessed: 29 March 2016, [Online]. Available: https://www.epa.gov/isa/integrated-science-assessment-isa-ozone.

[13]   Enivronmental Protection Agency (EPA). (2009). Integrated Science Assessment for Particulate Matter. 2015-03-20, [Online]. Available: https://www.epa.gov/isa/integrated-science-assessment-isa-particulate-matter.

[14]   A. Nel, T. Xia, L. Mädler, and N. Li, "Toxic potential of materials at the nanolevel", *Science*, vol. 311, no. 5761, pp. 622–627, 2006.

[15]   M. Auffan, J. Rose, J.-Y. Bottero, G. V. Lowry, J.-P. Jolivet, and M. R. Wiesner, "Towards a definition of inorganic nanoparticles from an environmental, health and safety perspective", *Nature nanotechnology*, vol. 4, no. 10, pp. 634–641, 2009.

[16]   M. Geiser, B. Rothen-Rutishauser, N. Kapp, S. Schürch, W. Kreyling, H. Schulz, M. Semmler, V. Im Hof, J. Heyder, and P. Gehr, "Ultrafine Particles Cross Cellular Membranes by Nonphagocytic Mechanisms in Lungs and in Cultured Cells", *Environmental Health Perspectives*, vol. 113, no. 11, pp. 1555–1560, 2005.

[17]   L. Ntziachristos, Z. Ning, M. D. Geller, and C. Sioutas, "Particle Concentration and Characteristics near a Major Freeway with Heavy-Duty Diesel Traffic", *Environmental Science & Technology*, vol. 41, no. 7, pp. 2223–2230, 2007.

[18]   Enivronmental Protection Agency (EPA). (2013). Integrated Science Assessment for Lead. Accessed: 29 March 2016, [Online]. Available: https://www.epa.gov/isa/integrated-science-assessment-isa-lead.

[19]   J. D. Marshall, E. Nethery, and M. Brauer, "Within-urban variability in ambient air pollution: Comparison of estimation methods", *Atmospheric Environment*, vol. 42, no. 6, pp. 1359–1369, 2008.

[20]   P. E. Benson, "A review of the development and application of the CALINE3 and 4 models", *Atmospheric Environment. Part B. Urban Atmosphere*, vol. 26, no. 3, pp. 379–390, 1992.

[21]   D. Oettl, "Documentation of the Lagrangian Particle Model Gral (Graz Lagrangian Model) Vs. 14.8", Amt der Steiermärkischen Landesregierung, Tech. Rep., 2014.

[22]   R. Liard, M. Zureik, Y. Le Moullec, D. Soussan, M. Glorian, A. Grimfeld, and F. Neukirch, "Use of personal passive samplers for measurement of NO 2, NO, and O 3 levels in panel studies", *Environmental research*, vol. 81, no. 4, pp. 339–348, 1999.

[23]   D. Kim, A. Sass-Kortsak, J. T. Purdham, R. E. Dales, and J. R. Brook, "Associations between personal exposures and fixed-site ambient measurements of fine particulate matter, nitrogen dioxide, and carbon monoxide in Toronto, Canada", *Journal of Exposure Science and Environmental Epidemiology*, vol. 16, no. 2, pp. 172–183, 2006.

[24] E. Palmes, A. Gunnison, J. DiMattio, and C. Tomczyk, "Personal sampler for nitrogen dioxide", *The American Industrial Hygiene Association Journal*, vol. 37, no. 10, pp. 570–577, 1976.

[25] Y. Yanagisawa and H. Nishimura, "A badge-type personal sampler for measurement of personal exposure to NO2 and NO in ambient air", *Environment International*, vol. 8, no. 1-6, pp. 235–242, 1982.

[26] Centre for Scientific Computing, University of Cambridge. (2011). Sensor Networks for Air Quality at Heathrow Airport. Accessed: 2 April 2016, [Online]. Available: http://www.snaq.org/.

[27] R. Camilli, C. M. Reddy, D. R. Yoerger, B. A. Van Mooy, M. V. Jakuba, J. C. Kinsey, C. P. McIntyre, S. P. Sylva, and J. V. Maloney, "Tracking hydrocarbon plume transport and biodegradation at Deepwater Horizon", *Science*, vol. 330, no. 6001, pp. 201–204, 2010.

[28] C. Stachniss, C. Plagemann, and A. J. Lilienthal, "Learning gas distribution models using sparse Gaussian process mixtures", *Autonomous Robots*, vol. 26, no. 2-3, pp. 187–202, 2009.

[29] P. P. Neumann, S. Asadi, A. J. Lilienthal, M. Bartholmai, and J. H. Schiller, "Autonomous Gas-Sensitive Microdrone: Wind Vector Estimation and Gas Distribution Mapping", *IEEE Robotics Automation Magazine*, vol. 19, no. 1, pp. 50–61, 2012.

[30] Imperial College London. (2006). MESSAGE - Mobile Environmental Sensing System Across Grid Environments. Accessed: 2 April 2016, [Online]. Available: http://www.commsp.ee.ic.ac.uk/~wiser/message/.

[31] Centre for Scientific Computing, University of Cambridge. (2011). Cambridge Mobile Urban Sensing (CamMobSens). Accessed: 2 April 2016, [Online]. Available: http://www.escience.cam.ac.uk/mobiledata/.

[32] M. Mead, O. Popoola, G. Stewart, P. Landshoff, M. Calleja, M. Hayes, J. Baldovi, M. McLeod, T. Hodgson, J. Dicks, A. Lewis, R. Cohen, J. annd Baron, J. Saffell, and R. Jones, "The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks", *Atmospheric Environment*, vol. 70, pp. 186–203, 2013.

[33] E. Bales, N. Nikzad, N. Quick, C. Ziftci, K. Patrick, and W. Griswold, "Citisense: Mobile air quality sensing for individuals and communities Design and deployment of the Citisense mobile air-quality system", in *The 6th International Conference on Pervasive Computing Technologies for Healthcare*, 2012, pp. 155–158.

[34] L. Sánchez, V. Gutiérrez, J. A. Galache, P. Sotres, J. R. Santana, J. Casanueva, and L. Muñoz, "SmartSantander: Experimentation and service provision in the smart city", in *16th International Symposium on Wireless Personal Multimedia Communications*, 2013, pp. 1–6.

[35] B. Elen, J. Theunis, S. Ingarra, A. Molino, J. Van den Bossche, M. Reggente, and V. Loreto, "The EveryAware SensorBox: a tool for community-based air quality monitoring", in *Sensing a Changing World Workshop*, 2012.

[36]   Preemptive Media. (2006). AIR - Area's Immediate Reading. Accessed: 3 April 2016, [Online]. Available: http://www.pm-air.net/index.php.

[37]   P. Dutta, P. M. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff, "Common Sense: Participatory Urban Sensing using a Network of Handheld Air Quality Monitors", in *Proceedings of the 7th ACM conference on embedded networked sensor systems*, 2009, pp. 349–350.

[38]   W. Willett, P. Aoki, N. Kumar, S. Subramanian, and A. Woodruff, "Common Sense Community: Scaffolding Mobile Sensing and Analysis for Novice Users", in *Pervasive Computing*, Springer, 2010, pp. 301–318.

[39]   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*, 2014, pp. 251–265.

[40]   SENSEable City Laboratory, MIT. (2009). The Copenhagen Wheel. Accessed: 3 April 2016, [Online]. Available: http://senseable.mit.edu/copenhagenwheel/.

[41]   B. Elen, J. Peters, M. V. Poppel, N. Bleux, J. Theunis, M. Reggente, and A. Standaert, "The Aeroflex: A Bicycle for Mobile Air Quality Measurements", *Sensors*, vol. 13, no. 1, pp. 221–240, 2012.

[42]   W. Hedgecock, P. Völgyesi, A. Ledeczi, X. Koutsoukos, A. Aldroubi, A. Szalay, and A. Terzis, "Mobile Air Pollution Monitoring Network", in *Proceedings of the 2010 ACM Symposium on Applied Computing*, 2010, pp. 795–796.

[43]   V. Sivaraman, J. Carrapetta, K. Hu, and B. G. Luxan, "Hazewatch: A participatory sensor system for monitoring air pollution in sydney", in *Local Computer Networks Workshops (LCN Workshops), 2013 IEEE 38th Conference on*, 2013, pp. 56–64.

[44]   S. Devarakonda, P. Sevusu, H. Liu, R. Liu, L. Iftode, and B. Nath, "Real-time air quality monitoring through mobile sensing in metropolitan areas", in *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, 2013, p. 15.

[45]   N. Castell, H.-Y. Liu, M. Kobernus, A. J. Berre, J. Noll, E. Cagatay, and R. Gangdal, "Mobile technologies and personalized environmental information for supporting sustainable mobility in Oslo: The Citi-Sense-MOB approach", in *28th EnviroInfo Conference*, 2014, pp. 699–706.

[46]   **A. Arfire**, A. Marjovi, A. Bahr, E. Droz, J. Eberle, J.-P. Calbimonte, K. Aberer, and A. Martinoli, "Sensing the Air We Breathe: Designing a Mobile Air Quality Monitoring Network", 2016, Working paper.

[47]   **A. Arfire**, A. Marjovi, and A. Martinoli, "Enhancing Measurement Quality through Active Sampling in Mobile Air Quality Monitoring Sensor Network", in *Proceedings of the 2016 IEEE International Conference on Advanced Intelligent Mechatronics*, 2016.

[48] **A. Arfire**, A. Marjovi, and A. Martinoli, "Mitigating Slow Dynamics of Low-Cost Chemical Sensors for Mobile Air Quality Monitoring Sensor Networks", in *Proceedings of the International Conference on Embedded Wireless Systems and Networks*, 2016, pp. 159–167.

[49] **A. Arfire**, A. Marjovi, and A. Martinoli, "Model-based Rendezvous Calibration of Mobile Sensor Networks for Monitoring Air Quality", in *Proceedings of the 2015 IEEE SENSORS Conference*, 2015, pp. 366–369.

[50] A. Marjovi, **A. Arfire**, and A. Martinoli, "Leveraging Land-use Data and Machine Learning for Extending Urban Air Pollution Maps Beyond the Coverage of a Mobile Sensor Network", submitted to the 14th ACM Conference on Embedded Networked Sensor Systems, 2016.

[51] A. Marjovi, **A. Arfire**, and A. Martinoli, "High Resolution Air Pollution Maps in Urban Environments using Mobile Sensor Networks", in *Proceedings of the 11th International Conference on Distributed Computing in Sensor Systems*, 2015, pp. 11–20.

[52] G. Bishop and D. Stedman, "Measuring the emissions of passing cars", *Accounts of Chemical Research*, vol. 29, no. 10, pp. 489–495, 1996.

[53] M. Aleixandre and M. Gerboles, "Review of small commercial sensors for indicative monitoring of ambient gas", *Chemical Engineering Transactions*, vol. 30, pp. 169–174, 2012.

[54] E. G. Snyder, T. H. Watkins, P. A. Solomon, E. D. Thoma, R. W. Williams, G. S. Hagler, D. Shelow, D. A. Hindin, V. J. Kilaru, and P. W. Preuss, "The changing paradigm of air pollution monitoring", *Environmental Science & Technology*, vol. 47, no. 20, pp. 11 369–11 377, 2013.

[55] SGX Sensortech S.A. (2014). MiCS-OZ-47 Sensor Datasheet. Accessed: 7 April 2016, [Online]. Available: http://www.sgxsensortech.com/content/uploads/2014/07/140905-SPF-2086-Data-Sheet-OZ-47.pdf.

[56] Aurora Scientific Inc. (2016). 200B miniPID Fast Response Gas Sensor. Accessed: 6 April 2016, [Online]. Available: http://hubspot.aurorascientific.com/product/fr-minipid/.

[57] City Technology Ltd. (2005). A3CO EnviroceL Sensor Datasheet. Accessed: 7 April 2016, [Online]. Available: https://www.citytech.com/PDF-Datasheets/a3co.pdf.

[58] KWJ Engineering, Inc. (2008). NO2-SNL Sensor Datasheet. Accessed: 7 April 2016, [Online]. Available: http://www.kwjengineering.com/wp-content/uploads/2008/05/no2_snl_0108.pdf.

[59] SGX Sensortech S.A. (2014). MiCS-2614 Sensor Datasheet. Accessed: 7 April 2016, [Online]. Available: http://www.sgxsensortech.com/content/uploads/2014/07/140905-SPF-2086-Data-Sheet-OZ-47.pdf.

[60] Sensirion. (2011). Datasheet SHT1x. Accessed: 7 April 2016, [Online]. Available: https://www.sensirion.com/fileadmin/user_upload/customers/sensirion/Dokumente/Humidity_Sensors/Sensirion_Humidity_Sensors_SHT1x_Datasheet_V5.pdf.

Bibliography

[61]   GE Measurement & Control Solutions. (2011). Telaire 6613 CO_2 Module. Accessed: 7 April 2016, [Online]. Available: http://www.ge-mcs.com/download/co2-flow/920-448G-LR.pdf.

[62]   TSI, Inc. (2014). Application note CPC-002. Accessed: 7 April 2016, [Online]. Available: http://www.tsi.com/uploadedFiles/_Site_Root/Products/Literature/Application_Notes/CPC-002-appnote.pdf.

[63]   Testo AG. (2015). DiSCmini – handheld particle counter Brochure. Accessed: 7 April 2016, [Online]. Available: http://testo-partikel.de/files/DiSCmini_EN_compressed.pdf.

[64]   Naneos Particle Solutions GmbH. (2015). Partector Aerosol Dosimeter - Operation Manual. Accessed: 7 April 2016, [Online]. Available: http://www.naneos.ch/pdf/naneos_partector_operation_manual.pdf.

[65]   D. Canvas. (2015). DIY Sensor Info. Accessed: 7 April 2016, [Online]. Available: http://datacanvas.org/sense-your-city/diy-sensor-info/.

[66]   M. Fierz, C. Houle, P. Steigmeier, and H. Burtscher, "Design, calibration, and field performance of a miniature diffusion size classifier", *Aerosol Science and Technology*, vol. 45, no. 1, pp. 1–10, 2011.

[67]   D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, T. Arn, J. Beutel, and L. Thiele, "Deriving high-resolution urban air pollution maps using mobile sensor nodes", *Pervasive and Mobile Computing*, vol. 16, pp. 268–285, 2015.

[68]   M. Fierz, D. Meier, P. Steigmeier, and H. Burtscher, "Aerosol measurement by induced currents", *Aerosol Science and Technology*, vol. 48, no. 4, pp. 350–357, 2014.

[69]   Microchip Technology Inc. (2005). dsPIC33F Family Data Sheet. Accessed: 8 April 2016, [Online]. Available: http://ww1.microchip.com/downloads/en/devicedoc/70165a.pdf.

[70]   u-blox AG. (2013). LEA-6R - u-blox 6 dead reckoning GPS module. Accessed: 9 April 2016, [Online]. Available: https://www.u-blox.com/sites/default/files/products/documents/LEA-6R_ProductSummary_(UBX-13003437).pdf.

[71]   Applanix. (2012). POS LV 120 Specifications. Accessed: 9 April 2016, [Online]. Available: http://www.applanix.com/media/POSLV%20Specs1012.pdf.

[72]   K. Aberer, M. Hauswirth, and A. Salehi, "Infrastructure for data processing in large-scale interconnected sensor networks", in *2007 International Conference on Mobile Data Management*, 2007, pp. 198–205.

[73]   N. Dawes, K. A. Kumar, S. Michel, K. Aberer, and M. Lehning, "Sensor metadata management and its application in collaborative environmental research", in *EScience, 2008. eScience'08. IEEE Fourth International Conference on*, IEEE, 2008, pp. 143–150.

[74]   J. Soldatos, N. Kefalakis, M. Hauswirth, M. Serrano, J.-P. Calbimonte, M. Riahi, K. Aberer, P. P. Jayaraman, A. Zaslavsky, I. P. Žarko, *et al.*, "OpenIoT: Open Source Internet-of-Things in the Cloud", in *Interoperability and Open-Source Solutions for the Internet of Things*, 2015, pp. 13–25.

[75]   J. Beutel, S. Gruber, S. Gubler, A. Hasler, M. Keller, R. Lim, I. Talzi, L. Thiele, C. Tschudin, and M. Yücel, "The PermaSense Remote Monitoring Infrastructure", in *International Snow Science Workshop*, vol. 9, 2009, pp. 187–191.

[76]   Commission Transports Lausanne Region. (2010). Comptages periodiques de trafic TI-TC 2010 - Rapport technique. Accessed: 9 April 2016, [Online]. Available: {http://www.lausanneregion.ch/documents/Synthese_TI_TC_2010final.pdf}.

[77]   J. J. Li, A. Jutzeler, and B. Faltings, "Estimating Urban Ultrafine Particle Distributions with Gaussian Process Models", *In S. Winter and C. Rizos (Eds.): Research@ Locate'14*, pp. 145–153, 2014.

[78]   D. Hasenfratz, O. Saukh, and L. Thiele, "On-the-fly calibration of low-cost gas sensors", *Wireless Sensor Networks*, pp. 228–244, 2012.

[79]   O. Saukh, D. Hasenfratz, C. Walser, and L. Thiele, "On rendezvous in mobile sensing networks", in *Real-World Wireless Sensor Networks*, 2014, pp. 29–42.

[80]   O. Saukh, D. Hasenfratz, and L. Thiele, "Reducing multi-hop calibration errors in large-scale mobile sensor networks", in *14th International Conference on Information Processing in Sensor Networks*, Seattle, WA, USA, 2015, pp. 274–285.

[81]   B. Maag, O. Saukh, D. Hasenfratz, and L. Thiele, "Pre-deployment testing, augmentation and calibration of cross-sensitive sensors", in *International Conference on Embedded Wireless Systems and Networks*, 2016, pp. 169–180.

[82]   A. Lilienthal and T. Duckett, "Building gas concentration gridmaps with a mobile robot", *Robotics and Autonomous Systems*, vol. 48, no. 1, pp. 3–16, 2004.

[83]   C. Stachniss, C. Plagemann, A. J. Lilienthal, and W. Burgard, "Gas distribution modeling using sparse gaussian process mixture models.", in *Robotics: Science and Systems IV*, 2009, pp. 310–317.

[84]   A. Kohnotoh and H. Ishida, "Active stereo olfactory sensing system for localization of gas/odor source", in *7th International Conference on Machine Learning and Applications*, 2008, pp. 476–481.

[85]   A. Lilienthal and T. Duckett, "A stereo electronic nose for a mobile inspection robot", in *1st International Workshop on Robotic Sensing*, 2003.

[86]   T. Lochmatter and A. Martinoli, "Tracking an odor plume in a laminar wind field with the crosswind-surge algorithm", in *Proceedings of the 11th International Symposium on Experimental Robotics*, vol. 54, 2008, pp. 473–482.

[87]   J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, Q. V. Le, *et al.*, "An inspection of signal dynamics using an open sampling system for gas identification", in *Workshop in Networked and Mobile Robot Olfaction in Natural, Dynamic Environments, IEEE International Conference on Robotics and Automation*, 2010.

[88]   Pea Soup Ltd. (2013). Wind Tunnel Air Flow Tracer Brochure. Accessed: 7 April 2016, [Online]. Available: http://www.smokemachines.net/PDF/Air-flow-tracer-brochure.pdf.

[89]   Sensidyne, Inc. (2001). A Plus DEEE All Specifications. Accessed: 29 May 2016, [Online]. Available: http://www.sensidyne.com/library/micro-air-pumps/A%20Plus%20DEEE%20All%20Specs.pdf.

[90]   Delta Electronics, Inc. (2008). AFB25x10mm-A Specifications. Accessed: 29 May 2016, [Online]. Available: http://partner.delta-corp.com/Products/FANUploads/Specification/AFB25x10mm-A.pdf.

[91]   M. C. Carotta, G. Martinelli, L. Crema, C. Malagù, M. Merli, G. Ghiotti, and E. Traversa, "Nanostructured thick-film gas sensors for atmospheric pollutant monitoring: quantitative analysis on field tests", *Sensors and Actuators B: Chemical*, vol. 76, no. 1, pp. 336–342, 2001.

[92]   M. Kamionka, P. Breuil, and C. Pijolat, "Calibration of a multivariate gas sensing device for atmospheric pollution measurement", *Sensors and Actuators B: Chemical*, vol. 118, no. 1, pp. 323–327, 2006.

[93]   A. Wijga, H. Smit, B. Brunekreef, J. Gerritsen, M. Kerkhof, L. Koopman, and H. Neijens, "Are children at high familial risk of developing allergy born into a low risk environment? the PIAMA Birth Cohort Study", *Clinical & Experimental Allergy*, vol. 31, no. 4, pp. 576–581, 2001.

[94]   B. Ponte, M. Pruijm, P. Marques-Vidal, P.-Y. Martin, M. Burnier, F. Paccaud, G. Waeber, P. Vollenweider, and M. Bochud, "Determinants and burden of chronic kidney disease in the population-based CoLaus study: a cross-sectional analysis", *Nephrology Dialysis Transplantation*, pp. 1–11, 2013.

[95]   H. Alwan, M. Pruijm, B. Ponte, D. Ackermann, I. Guessous, G. Ehret, J. A. Staessen, K. Asayama, P. Vuistiner, S. E. Younes, *et al.*, "Epidemiology of masked and white-coat hypertension: the family-based SKIPOGH study", *PloS one*, vol. 9, no. 3, pp. 1–9, 2014.

[96]   A. Jakeman, R. Simpson, and J. Taylor, "Modeling distributions of air pollutant concentrations - III. The hybrid deterministic-statistical distribution approach", *Atmospheric Environment*, vol. 22, no. 1, pp. 163–174, 1988.

[97]   D. W. Wong, L. Yuan, and S. A. Perlin, "Comparison of spatial interpolation methods for the estimation of air quality data", *Journal of Exposure Science and Environmental Epidemiology*, vol. 14, no. 5, pp. 404–415, 2004.

[98]   E. G. Dragomir, "Air quality index prediction using K-nearest neighbor technique", *Bulletin of PG University of Ploiesti, Series Mathematics, Informatics, Physics, LXII*, vol. 1, no. 2010, pp. 103–108, 2010.

[99]   H. Niska, T. Hiltunen, A. Karppinen, J. Ruuskanen, and M. Kolehmainen, "Evolving the neural network model for forecasting air pollution time series", *Engineering Applications of Artificial Intelligence*, vol. 17, no. 2, pp. 159–167, 2004.

[100] J. Yi and V. R. Prybutok, "A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area", *Environmental Pollution*, vol. 92, no. 3, pp. 349–357, 1996.

[101] T. Hussein, A. Karppinen, J. Kukkonen, J. Härkönen, P. P. Aalto, K. Hämeri, V.-M. Kerminen, and M. Kulmala, "Meteorological dependence of size-fractionated number concentrations of urban aerosol particles", *Atmospheric Environment*, vol. 40, no. 8, pp. 1427–1440, 2006.

[102] B. Mølgaard, T. Hussein, J. Corander, and K. Hämeri, "Forecasting size-fractionated particle number concentrations in the urban atmosphere", *Atmospheric Environment*, vol. 46, pp. 155–163, 2012.

[103] S. Clifford, S. L. Choy, T. Hussein, K. Mengersen, and L. Morawska, "Using the generalised additive model to model the particle number count of ultrafine particles", *Atmospheric Environment*, vol. 45, no. 32, pp. 5934–5945, 2011.

[104] M. Reggente, J. Peters, J. Theunis, M. Van Poppel, M. Rademaker, P. Kumar, and B. De Baets, "Prediction of ultrafine particle number concentrations in urban environments by means of Gaussian process regression based on measurements of oxides of nitrogen", *Environmental Modelling & Software*, vol. 61, pp. 135–150, 2014.

[105] X. Hu, L. A. Waller, M. Z. Al-Hamdan, W. L. Crosson, M. G. Estes, S. M. Estes, D. A. Quattrochi, J. A. Sarnat, and Y. Liu, "Estimating ground-level PM 2.5 concentrations in the southeastern US using geographically weighted regression", *Environmental research*, vol. 121, pp. 1–10, 2013.

[106] A. Jutzeler, J. J. Li, and B. Faltings, "A region-based model for estimating urban air pollution", in *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[107] M. Haklay and P. Weber, "OpenStreetMap: User-generated street maps", *Pervasive Computing, IEEE*, vol. 7, no. 4, pp. 12–18, 2008.

[108] N. Schuessler and K. W. Axhausen, "Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT)", *Working paper 568, Institute for Transport Planning and System (IVT)*, Oct. 2009.

[109] M. Zeri, J. F. Oliveira-Júnior, and G. B. Lyra, "Spatiotemporal analysis of particulate matter, sulfur dioxide and carbon monoxide concentrations over the city of Rio de Janeiro, Brazil", *Meteorology and Atmospheric Physics*, vol. 113, no. 3-4, pp. 139–152, 2011.

[110] E. Limpert, W. A. Stahel, and M. Abbt, "Log-normal distributions across the sciences: keys and clues", *BioScience*, vol. 51, no. 5, 2001.

[111] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012, vol. 936.

[112] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[113] Y. Bengio, "Learning deep architectures for ai", *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[114]   G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[115]   H. Lee, C. Ekanadham, and A. Y. Ng, "Sparse deep belief net model for visual area V2", in *Advances in neural information processing systems*, 2008, pp. 873–880.

[116]   P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion", *The Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.

[117]   J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization", *The Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2011.

[118]   J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, Q. V. Le, *et al.*, "Large scale distributed deep networks", in *Advances in Neural Information Processing Systems*, 2012, pp. 1223–1231.

[119]   M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems", *White paper, Google Research*, 2015.

Curriculum Vitae

Adrian Arfire

Education

2010-2016	PhD in Robotics, Control, and Intelligent Systems *Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland*
2008-2010	MS in Microengineering *Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland*
2003-2008	Dipl.-Ing. in Systems and Computers *"Politehnica" University of Bucharest, Romania*

Experience

2009-2010	Research Assistant *Distributed Intelligent Systems and Algorithms Laboratory, EPFL, Switzerland*
2009-2010	Research Assistant *Chair of International Finance, EPFL, Switzerland*
2009	Student Teaching Assistant *School of Engineering, EPFL, Switzerland*
2007-2008	Student Intern *Infineon Technologies Romania & CO SCS, Bucharest, Romania*

Publications and Working Papers

1. **A. Arfire**, A. Marjovi, A. Bahr, E. Droz, J. Eberle, J.-P. Calbimonte, K. Aberer, and A. Martinoli, "Sensing the Air We Breathe: Designing a Mobile Air Quality Monitoring Network", 2016, Working paper

2. A. Marjovi, **A. Arfire**, and A. Martinoli, "Leveraging Land-use Data and Machine Learning for Extending Urban Air Pollution Maps Beyond the Coverage of a Mobile Sensor Network",

submitted to the 14th ACM Conference on Embedded Networked Sensor Systems, 2016

3. **A. Arfire**, A. Marjovi, and A. Martinoli, "Enhancing Measurement Quality through Active Sampling in Mobile Air Quality Monitoring Sensor Network", in *Proceedings of the 2016 IEEE International Conference on Advanced Intelligent Mechatronics*, 2016

4. **A. Arfire**, A. Marjovi, and A. Martinoli, "Mitigating Slow Dynamics of Low-Cost Chemical Sensors for Mobile Air Quality Monitoring Sensor Networks", in *Proceedings of the International Conference on Embedded Wireless Systems and Networks*, 2016, pp. 159–167

5. **A. Arfire**, A. Marjovi, and A. Martinoli, "Model-based Rendezvous Calibration of Mobile Sensor Networks for Monitoring Air Quality", in *Proceedings of the 2015 IEEE SENSORS Conference*, 2015, pp. 366–369

6. A. Marjovi, **A. Arfire**, and A. Martinoli, "High Resolution Air Pollution Maps in Urban Environments using Mobile Sensor Networks", in *Proceedings of the 11th International Conference on Distributed Computing in Sensor Systems*, 2015, pp. 11–20

7. A. Prorok, **A. Arfire**, A. Bahr, J. R. Farserotu, and A. Martinoli, "Indoor navigation research with the Khepera III mobile robot: An experimental baseline with a case-study on ultra-wideband positioning", in *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2010, pp. 1–9

Invited Talks

1. "Enabling High Resolution Urban Pollution Monitoring through Mobile Sensor Networks", Second International Workshop on New Sensing Technologies for Indoor and Outdoor Air Quality Control, Brindisi, Italy, March 2014

Project Supervision

1. Francesco Giardina and Camille Fallous, Semester Project (2016)
   Developing a Probabilistic Graphical Model for Fine Particle Estimation using OpenSense and Land-use Data

2. Fanny Jeanneret-Grosjean, Master's Thesis (2015)
   Sensing the Atmospheric Ozone from Space and *In Situ* Measurements – A Comparative Analysis

3. Fabrizzio Gonzales, Semester Project (2015)
   Generating Air Pollution Maps in Lausanne using the OpenSense Fine Particle Network Data

4. Fanny Jeanneret-Grosjean, Semester Project (2015)
   Generating Lausanne Pollution Maps using the OpenSense Network

5. Patrick Osterwalder, Semester Project (2014)
   Generating Lausanne Pollution Maps using the OpenSense Network

6. Thomas Coral and Jonathan Giezendanner, Semester Project (2014)
   Fine Particle Distribution Estimation in Lausanne using the OpenSense Network

7. Nicolas Bigler, Civil Service Project (2013)
   OpenSense – Data Validation and Sensor Calibration

8. Kim Ehrensperger and Laurent Hafiz, Semester Project (2013)
   Developing an Air Quality Layer for an Integrated Mobile Sensor Network Simulator

9. Loïc Frund, Civil Service Project (2012)
   Integrated Simulation Framework for the OpenSense Project

10. Christophe Paccolat, Civil Service Project (2012)
    Intelligent Air Sampling System for a Vehicular Wireless Sensor Network

11. Sébastien Martineau, Summer Internship (2012)
    Enhanced Outdoor Localization for an Urban Mobile Sensor Network

12. Kévin André and Alex Bonvin, Semester Project (2012)
    Gas Field Estimation - A Challenge for Mobile Wireless Sensor Networks

13. Benjamin Fankhauser, Civil Service Project (2012)
    Using Quadrotors to Monitor the Environment

Academic Service (Reviews)

1. International Conference on Advanced Intelligent Mechatronics (AIM), 2016

2. International Symposium on Distributed Autonomous Robotic Systems (DARS), 2012

3. International Conference on Intelligent Robots and Systems (IROS), 2011

Languages

Romanian	native
English	fluent
French	fluent
Italian	conversational

Personal Details

Date of birth:	10 June 1984
Place of birth:	Galaţi, Romania
Citizenship:	Romania