

DISS. ETH NO. 00000

Enabling Large-Scale Urban Air Quality Monitoring with Mobile Sensor Nodes

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
DAVID HASENFRATZ
M.Sc. ETH Zurich

born on 26.11.1984
citizen of Switzerland, Hungary

accepted on the recommendation of
Prof. Dr. Lothar Thiele, examiner
Prof. Dr. Koen Langendoen, co-examiner

2015



Institut für Technische Informatik und Kommunikationsnetze
Computer Engineering and Networks Laboratory

TIK-SCHRIFTENREIHE NR. 000

David Hasenfratz

Enabling Large-Scale Urban Air Quality Monitoring with Mobile Sensor Nodes



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

A dissertation submitted to
ETH Zurich
for the degree of Doctor of Sciences

DISS. ETH NO. 00000

Prof. Dr. Lothar Thiele, examiner
Prof. Dr. Koen Langendoen, co-examiner

Examination date: January 28, 2015

Abstract

Urban air pollution is a major concern in many cities worldwide. Atmospheric pollutants are responsible for health problems ranging from asthma to cancer. Air pollution also causes environmental damages.

Monitoring airborne pollutants is of utmost importance to reliably assess the impact of air pollution on the human health, enable urban planners to craft and accurately evaluate new policies, and increase public awareness. Nowadays, air pollution is monitored by networks of highly accurate but fixed measurement stations. Hence, the gathered data has a low spatial resolution and can not be used to assess the spatial variability of pollutants in detail. As a result, little is known about the spatial distribution of air pollutants in urban environments.

In this thesis, we tackle this challenge and derive fine-grained intraurban pollution maps valuable for a range of applications. We use compact low-cost sensors installed on top of public transport vehicles to obtain a high spatial measurement resolution within a large urban area. We develop algorithms, which allow us to accurately monitor the phenomena of interest despite using noisy, low-cost sensors. Finally, we use the measurement to derive air pollution maps with a high spatial and temporal resolution. The main contributions of this thesis are:

- We build a mobile air quality monitoring network by equipping public transport vehicles with low-cost air quality sensor nodes collecting spatially resolved measurements. It is the first mobile air pollution monitoring network operating for over three years by now.
- We are the first to study multi-hop calibration of mobile sensor networks, with respect to a reference signal, in detail. We develop a new calibration algorithm to accurately calibrate networks of low-cost sensors by highly reducing error propagation in the network. Further, we assess the quality of the measurements by integrating generic models for the phenomena monitored and the sensors used.
- We describe a new modeling approach and use the measurements collected with our mobile network to derive accurate urban pollution maps with an unprecedented spatial and temporal resolution. The maps open up many new application opportunities. For example, we introduce a novel route planning service, which helps urban dwellers to reduce their exposure to airborne pollutants.

Zusammenfassung

Die Luftverschmutzung in städtischen Ballungsgebieten ist weltweit ein grosses Problem. Luftschadstoffe verursachen unterschiedlichste Krankheiten (hauptsächlich Erkrankungen der Atemwege und des Kreislaufsystems) und führen zu Umweltproblemen.

Deshalb ist die Kontrolle der Immissionsbelastung äusserst wichtig und wird heutzutage in vielen Ländern staatlich reguliert. Üblicherweise werden die Luftschadstoffe mit statischen Messnetzen überwacht, welche mit sehr genauen Messgeräten bestückt sind, aber durch die geringe Anzahl Messstationen keine flächendeckende Messabdeckung ermöglichen. Dies hat zur Folge, dass heute über die räumliche Verteilung der Luftschadstoffe innerhalb urbaner Gebiete wenig bekannt ist.

In dieser Arbeit stellen wir uns dieser Herausforderung und ermöglichen die Entwicklung von innerstädtischen Karten, welche detailliert die Belastung durch Luftschadstoffe aufzeigen können. Wir installieren kompakte und kostengünstige Sensoren auf Dächern von Trams und messen damit mit einer sehr hohen Auflösung die Luftverschmutzung entlang der Zürcher Tramlinien. Dazu entwerfen wir intelligente Algorithmen, welche uns erlauben eine gute Messgenauigkeit zu erreichen, ungeachtet der kleinen Grösse und dem günstigen Preis der eingebauten Sensoren. Dies ermöglicht uns Karten zu berechnen, die mit einer sehr hohen zeitlichen und räumlichen Auflösung, die Luftverschmutzung in der Stadt Zürich aufzeigen. Die Hauptbeiträge dieser Arbeit können folgenderweise zusammengefasst werden:

- Wir entwickeln ein einzigartiges mobiles Messnetz, welches mit schuhschachtelgrossen Boxen, installiert auf Dächern von Zürcher Trams, die Konzentration verschiedener Luftschadstoffe misst.
- Wir sind die Ersten, die eingehend die Multihop-Kalibrierung von mobilen Sensornetzwerken analysieren. Wir beschreiben ein neuartiges Verfahren, das die Fehlerfortpflanzung in Sensornetzwerken deutlich reduziert und dadurch eine viel genauere Sensorkalibrierung ermöglicht als bisher bekannte Verfahren.
- Wir benutzen ein ausgeklügeltes Modellierungsverfahren um, basierend auf unseren Daten, hochauflösende Karten zur Luftverschmutzung zu erstellen. Diese Karten erlauben neue Einsichten und ermöglichen eine grosse Zahl neuer Applikationen.

Acknowledgments

...tbd...

The work presented in this thesis was funded by NanoTera.ch with Swiss Confederation financing under the projects OpenSense and OpenSense II. This support is gratefully acknowledged.

Contents

Abstract	i
Zusammenfassung	iii
Acknowledgments	v
List of Figures	ix
List of Tables	xi
1 Introduction	1
1.1 State-of-the-art Air Quality Monitoring	2
1.2 Mobile Air Quality Monitoring	3
1.3 Challenges in Mobile Sensor Networks	6
1.4 Thesis Contributions	10
1.5 Thesis Outline	10
2 Mobile Air Pollution Monitoring Network	15
2.1 Long-Term Sensor Testbed	16
2.2 System Architecture	19
2.3 Deployments and Collected Data Sets	25
2.4 Related Work	31
2.5 Conclusions	33
3 Calibrating a Mobile Network of Low-Cost Sensors	35
3.1 Introduction	36
3.2 Assumptions and Models	39
3.3 Rendezvous Between Sensors	40
3.4 Calibration Methods	42
3.5 Reducing Error Accumulation	46
3.6 Network Calibration	55
3.7 Calibrating the Mobile Sensor Network	57
3.8 Related Work	60
3.9 Conclusions	64
4 Accuracy Bounds for Low-Cost Sensor Readings	67
4.1 Introduction	68
4.2 System Models	69
4.3 Computation of the Signal Range	73

4.4	Case Study	78
4.5	Evaluation	80
4.6	Related Work	89
4.7	Conclusions	90
5	Developing Fine-Grained Air Pollution Maps	93
5.1	Introduction	94
5.2	Spatially Highly Resolved Ultrafine Particle (UFP) Measurements	96
5.3	Data Validation	100
5.4	Developing Land-Use Regression Models to Create High-Resolution Pollution Maps	102
5.5	Revealing the Temporal Resolution Limit	107
5.6	Increasing the Temporal Resolution Limit	111
5.7	Related Work	116
5.8	Conclusions	117
6	The Health-Optimal Route Planner	119
6.1	Introduction	119
6.2	Road Network Graph	120
6.3	Finding the Least-Cost Path	121
6.4	Comparing Shortest Paths to Health-Optimal Paths	123
6.5	Smartphone Application	124
6.6	Related Work	127
6.7	Conclusions	128
7	Conclusions and Outlook	129
7.1	Contributions	129
7.2	Possible Future Directions	131
A	Monitoring Radio-Frequency Electromagnetic Fields	135
A.1	Introduction	135
A.2	Background	138
A.3	Electric Field Monitoring	139
A.4	Calibration and Validation	140
A.5	Measurement Campaign	143
A.6	Related Work	145
A.7	Conclusions	147
	Bibliography	149
	List of Publications	167
	Curriculum Vitæ	169

List of Figures

1.1	Regional-scale air pollution map from NABEL	3
1.2	Trade-offs for different system architecture classes	4
1.3	High-resolution pollution maps using mobile sensor nodes.	5
1.4	Semiconductor ozone (O_3) sensor	7
1.5	Electrochemical carbon monoxide (CO) sensor	7
1.6	Organization of this thesis.	11
2.1	Sensor testbed installation	16
2.2	Measurements from factory calibrated ozone sensors	17
2.3	Effect of sensor mobility on the sensor readings	19
2.4	Mobile air quality sensor node	21
2.5	Air intakes are equipped with protection covers	22
2.6	Current draw of the ozone sensor and USB translator	23
2.7	GasMobile hardware architecture and Android application	24
2.8	Air quality sensor nodes installed on top of streetcars	26
2.9	Average operation times of the mobile sensor nodes	27
2.10	Data latency of the streetcar deployment	28
2.11	Air quality sensor node is installed on top of the LuftiBus vehicle	29
2.12	GasMobile pollution measurements from several bicycle rides	30
3.1	The fluctuation of a low-cost sensor's optimal calibration curve	36
3.2	Linear dependence between raw sensor readings and phenomenon	39
3.3	The residuals minimized by <i>OLS</i> and <i>GMR</i>	43
3.4	Regression lines of <i>OLS</i> and <i>GMR</i>	45
3.5	Calibration path $u \rightarrow a \rightarrow \dots \rightarrow x \rightarrow y \rightarrow z$	47
3.6	Impact of measurement quantity and sensor noise on the variance estimate	50
3.7	Sensor calibration under fixed phenomenon range	51
3.8	Sensor calibration under high noise	52
3.9	Sensor calibration under variable phenomenon range	53
3.10	Sensor calibration under small phenomenon range	54
3.11	Sensor calibration under large phenomenon range	54
3.12	Comparison of <i>OLS</i> and <i>GMR</i> on a 35-node grid topology	57
3.13	Meeting points among mobile sensors and reference stations	58
3.14	Number of calibration paths with two to five sensor nodes.	59

3.15	Dependency between calibration error and calibration path length	61
4.1	System overview for computing individual accuracy bounds	70
4.2	Phenomenon model	72
4.3	The maximum sensor drift bounds the sensor accuracy	74
4.4	Erroneous measurements are detected with a conflict graph	77
4.5	The phenomenon gradient for temperature and ozone	79
4.6	Offset adjustment of the measurements	80
4.7	Temperature signal range sizes for different model settings	82
4.8	Ozone signal range sizes for different model settings	82
4.9	Signal ranges for static temperature and ozone measurements	83
4.10	Accuracy of the computed signal ranges	84
4.11	Detection of sensor failures	86
4.12	Dependence on the number of available reference measurements	87
4.13	Long-term characteristic of low-cost sensors	88
4.14	Signal ranges for mobile temperature and ozone measurements	89
5.1	Novel UFP concentration maps for Zurich, Switzerland	95
5.2	Number of UFP measurements over the period of two years	98
5.3	Ten mobile sensor nodes cover a large urban area	99
5.4	The log-normal distribution accurately fits the distribution of the measurements	100
5.5	Measurements from mobile low-cost sensors and fixed stations	102
5.6	Typical influence factors of explanatory variables	104
5.7	Seasonal UFP pollution maps with a resolution of 100 m × 100 m	106
5.8	Model predicted versus real measured UFP concentrations	108
5.9	Model performance for different temporal resolutions	110
5.10	Yearly pollution maps modeled with 20–1200 grid cells	111
5.11	Dependence between number of measurements and RSE	112
5.12	Novel modeling approach with history database	113
5.13	Accuracy of the history database	114
5.14	Quality of semi-daily pollution maps with history database	115
6.1	Road network graph for the computation of least-cost routes	120
6.2	Weight function for road segments	121
6.3	The A* pathfinding algorithm	122
6.4	Shortest and health-optimal routes of three exemplary source-destination pairs	123
6.5	Length and exposure differences between different paths	125
6.6	Application for iPhone and iPad to compute health-optimal paths	126
6.7	Path computation times on three smartphones and one tablet	127
A.1	Electromagnetic exposure levels in the city center of Zurich	137
A.2	Electromagnetic spectrum from 1 Hz to 1 PHz	138

A.3	Calibration of the RF-EMF measurements	142
A.4	GSM exposure map of Zurich, Switzerland	144
A.5	Dependence between explanatory variables and field strength	145
A.6	State-of-the-art RF-EMF exposure meters	146

List of Tables

2.1	Data sets from the sensor testbed installation	18
2.2	Data sets from the streetcar deployment	27
2.3	Data sets collected on top of the LuftiBus vehicle.	28
3.1	Evaluated data sets of the mobile air pollution monitoring network	58
3.2	Average calibration errors	59
3.3	Average number of sensors calibrated over a 10-day period . . .	60
5.1	The 12 explanatory variables examined to build pollution maps	103
A.1	Average calibration errors of the RF-EMF measurements	142
A.2	Mean field strengths in Zurich and five other European cities . .	143

1

Introduction

Air pollution is a major concern in many cities worldwide. Atmospheric pollutants considerably affect human health; they are responsible for a variety of respiratory (e.g., asthma) and cardiovascular diseases (e.g., heart attack) and are known to cause cancer if humans are exposed to them for extended periods of time [MOP⁺96]. According to the 2014 WHO report, in 2012 air pollution caused more than 7 million deaths worldwide [WHO14]:

WHO reports that in 2012 around 7 million people died—one in eight of total global deaths—as a result of air pollution exposure. This finding more than doubles previous estimates and confirms that air pollution is now the world’s largest single environmental health risk. Reducing air pollution could save millions of lives.

Additionally, air pollution is responsible for environmental problems, such as eutrophication and acidification of ecosystems. Hence, air pollution monitoring is of utmost importance to reliably assess the impact of air pollution on the human health and the environment. Better assessment of human exposure to air pollution, more knowledge about the diseases caused by air pollution, and the use of improved measurement technology enable better estimations about the impact of air pollution on our daily lives [WHO14].

At the present time, the concentration of air pollutants is mainly assessed by the measurements of fixed stations. These stations are highly reliable and able to accurately measure a wide range of air pollutants. However, their high acquisition and maintenance costs severely limit the

number of installations [CCG07]. As a result, very little is known about the spatial distribution of air pollutants in urban environments. There is a lack of reliable, intraurban air pollution maps, despite the need for it, especially for pollutants with high spatial variability. Such maps could raise the citizens' awareness about air pollution, enable environmental scientists to craft and evaluate new policies, and help epidemiologists to better quantify the impact of air pollution on the human health.

In this thesis, we introduce a mobile measurement network to monitor various air pollutants with a high spatial resolution. We develop algorithms to calibrate unstable and noisy low-cost gas sensors and to assess the reliability of the measurements. Then, we use the measurements to derive pollution maps with an accuracy that is comparable to state-of-the-art air pollution maps [HBdH⁺08], while achieving unprecedented spatio-temporal resolution. Subsequently, we use these pollution maps to create a smartphone application, which empowers city dwellers to assess and reduce their exposure to air pollution.¹

1.1 State-of-the-art Air Quality Monitoring

Nowadays, air pollution is monitored by networks of fixed measurement stations operated by official authorities. These fixed stations are highly reliable and able to accurately measure a wide range of environmental parameters (e.g., wind speed, solar radiation, precipitation) and air pollutants using traditional analytical instruments, such as mass spectrometers and gas chromatographs [Fed13]. These static monitoring networks are crucial to monitor all important airborne pollutants with a very high accuracy and temporal resolution. The measurements from these certified instruments provide regulators a solid basis to assess the air quality and allow them to be responsive and take counteractions in case of high exposure levels.

The drawbacks of these complex measurement systems are their large size, high price, and laborious maintenance. To assure high precision, the instruments are manually calibrated (adjusted in the parts-per-thousand range) every 14 days [Fed13]. During calibration, a sensor is exposed to a predefined set of gas concentrations, and the sensor's calibration parameters are adjusted such that the deviations between applied gas concentrations and sensor outputs are minimized. The extensive cost of acquiring and operating these stations severely

¹hRouting–The Health-Optimal Route Planner: <https://itunes.apple.com/us/app/hrouting-health-optimal-route/id931299863>

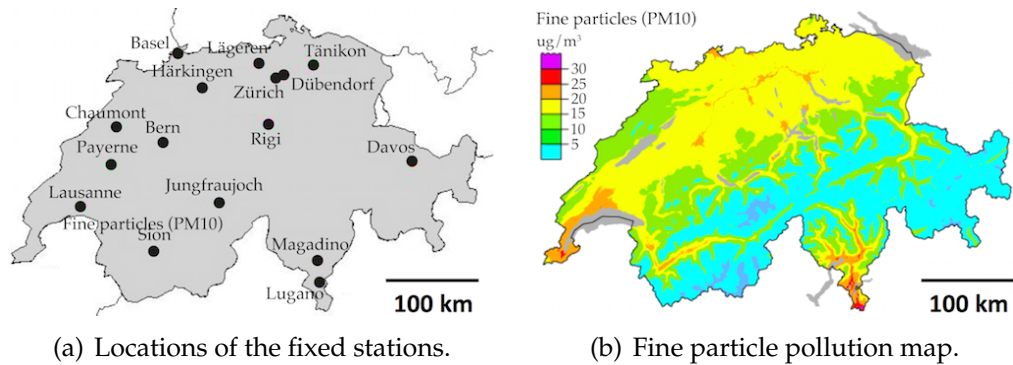


Figure 1.1: Measurements from the Swiss National Air Pollution Monitoring Network (NABEL) are used to derive regional-scale air pollution maps.

limits the number of installations [CCG07]. For example, the Swiss National Air Pollution Monitoring Network (NABEL) [Nat14a] operates 16 measurement stations, distributed over an area of 41,300 km², as depicted in Figure 1.1(a). The 16 stations monitor the pollution levels in seven typical exposure areas (e.g., urban heavy traffic, suburban, and rural). The whole area of Switzerland is also categorized in these seven areas and the measurements from the fixed stations are used to generate pollution maps, such as the one shown in Figure 1.1(b). These maps can be used to analyze regional differences but their spatial resolution is not high enough to derive intra-urban pollution distributions. It is important to notice that NABEL's measurement density is fairly high compared to other national monitoring networks [Nyf01].

The concentration of air pollutants is highly location-dependent. Traffic junctions, urban canyons, industrial installations, and topological structure all have considerable impact on local air pollution [VFPGF03]. Hence, in recent years several research groups started experimenting with mobile air pollution monitoring systems [HBPW08, VNKL08, DAK⁺09, CKCH09, DSL⁺13, RPV14].

1.2 Mobile Air Quality Monitoring

Mobile measurement systems trade off temporal coverage against spatial coverage, enabling a high spatial resolution across large areas without the need for a huge number of fixed nodes, e.g., the GreenOrbs deployment needs up to 330 fixed sensor nodes to realize an all-year ecological surveillance of a small forest area [MHL⁺09, LHL⁺11]. Mobile measurement networks often leverage existing vehicle-based networks (e.g., taxicab fleets [ALPR12], bicycle messengers [Kan10], and

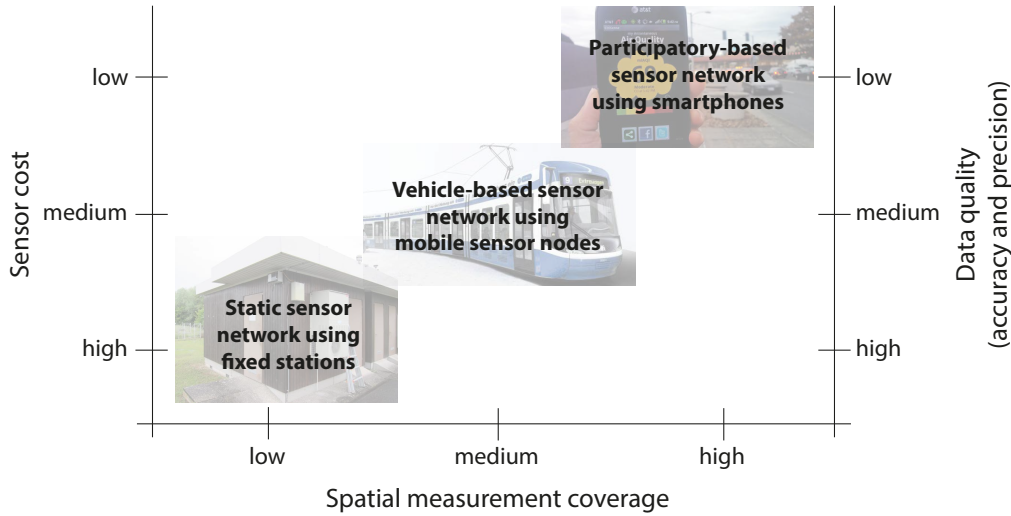


Figure 1.2: Trade-off between obtainable measurement coverage, tolerable sensor cost, and expected data quality for different system architecture classes.

public transport networks [RPV14]) and involve individuals from the general public to collect measurements [FOC⁺11]. In vehicle-based and participatory sensor networks it is not possible to use the same high-end sensing instruments as the ones deployed in fixed stations. Hence, with these types of system architecture we face a new interesting trade-off between obtainable measurement coverage and expected data quality, as depicted in Figure 1.2. In this thesis, we use measurements from static networks to enhance the quality of the measurements collected by vehicle-based and participatory-based sensor networks.

Vehicle-based sensor networks. The deployment of sensor nodes on top of mobile vehicles with either fixed (e.g., buses and streetcars of the public transport network [DSL⁺13, RPV14]) or arbitrary (e.g., private cars and taxicabs [VNKL08, CZZ⁺13]) movement patterns, enables the automatic collection of large, spatially resolved data sets. Especially, installing sensor nodes on top of public transport vehicles facilitates the coverage of a large urban area on a regular schedule to derive fine-grained pollution maps, as exemplified in Figure 1.3. While in vehicle-based sensor networks energy and size requirements are not as strict as in participatory sensing scenarios, there are still limitations on the available space, the acceptable weight, and the maximum allowed power consumption. This prohibits the installations of most standard instruments used in fixed stations. Further limitations are introduced as the sensor nodes are potentially in operation around the clock with limited maintenance possibilities and under harsh deployment settings.



(a) Sensor node on top of a public transport vehicle. (b) Fine-grained pollution map.

Figure 1.3: Mobile sensor nodes achieve a good measurement coverage, which is necessary to derive pollution maps with a high spatial resolution.

Participatory air quality monitoring. Spatially fine-grained air pollution data can also be acquired with a community-driven sensing infrastructure [BEH⁺06]. Such initiatives that pursue the public gathering of reliable data, gained increasing popularity in the last years, e.g., worldwide data collection of local food conditions and nuclear radiation.² These examples show that it is possible to collect region-wide measurements by involving the general public. Given the broad availability of personal GPS-equipped smartphones, these devices can be used to build a large-scale sensor network of mobile devices for participatory monitoring of pollution data [SD10, ZLW⁺14]. Involving the average citizen in sensing the air she breathes also helps to rise public awareness and encourages to move towards a sustainable development [UNO92].

However, obtaining and retaining the critical mass of participants to get a coherent picture of the exposure situation in an area of interest is a formidable challenge and may require hundreds to thousands of contributors. This is particularly difficult as long as smartphones are not equipped by default with the sensors required.

Combination of participatory and vehicle-based monitoring. In this thesis, we introduce a prototypical participatory air quality monitoring platform using an off-the-shelf smartphone, which allows anyone to collect air quality data. Additionally, we install sensor nodes equipped with multiple low-cost gas sensors on top of several public transport vehicles to achieve a good coverage of the whole city while only using a small number of nodes. The public transport network forms an attractive backbone for performing periodic measurements as their vehicles are

²<https://costofchicken.crowdmap.com>, <https://radiation.crowdmap.com>

operating *(i)* on a large number of spatially distributed and predefined routes, *(ii)* on a regular schedule, and *(iii)* with a high reliability. In the following, we describe the arising challenges in the use of mobile sensor nodes for urban air pollution monitoring.

1.3 Challenges in Mobile Sensor Networks

Operating mobile sensor networks and making use of data collected by these networks entail several challenges, which we discuss below.

1.3.1 Measurement System

The components used in mobile measurement systems have to be suitable for mobile measurements. Especially, the sensors need to have a short response time to deliver accurate measurements even while moving. Furthermore, typically mobile systems have limited processing power, memory storage, and upload bandwidth requiring data aggregation and data filtering techniques. Participatory and vehicle-based sensing system put forward some specific additional challenges.

Vehicle-based sensing systems. Sensor nodes installed on top of vehicles (e.g., taxicabs, buses, and streetcars) have usually relaxed requirements concerning form factor and power requirements. This simplifies part of the design of such sensor nodes, e.g., larger and more reliable components can be used compared to participatory sensing devices. However, nodes installed on top of a network of mobile vehicles have to endure harsh deployment settings. We have to expect constant vibrations, high temperature and humidity variations, and long, unattended operating times (often requiring a design for 24/7 operation).

Participatory-based sensing systems. Getting the general public involved in participatory air quality monitoring to collect useful data posts several challenges. These involve providing the user with an affordable low-cost and low-power measurement hardware with minimal maintenance requirements. The data acquisition and processing software on the user's device needs to be unobtrusive and user-friendly to not interfere with other applications in use. Furthermore, the data acquisition application should support the user in gathering high-quality data, e.g., by indicating and aiding sensor calibration. To motivate the user to keep up data collection, we are in need of a suitable information feedback system offering rewards and incentives to the participating users.

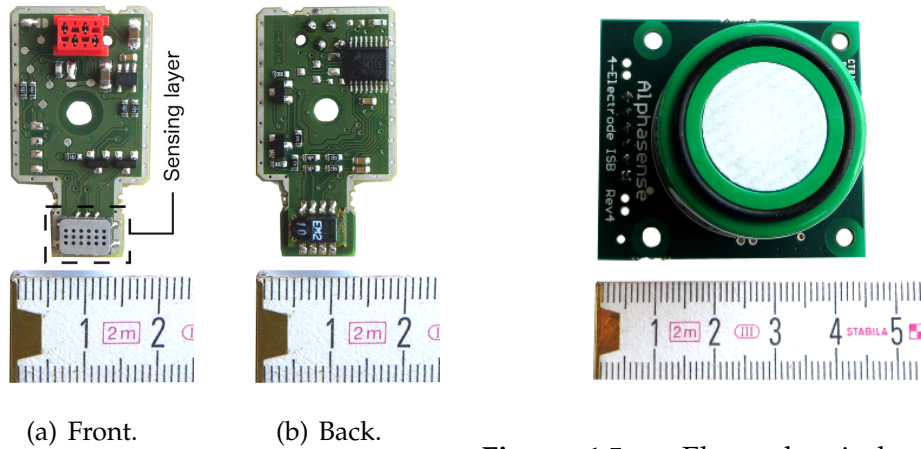


Figure 1.4: Semiconductor ozone (O_3) sensor with attached microprocessor.

Figure 1.5: Electrochemical carbon monoxide (CO) sensor with attached microprocessor transmitter board.

1.3.2 Sensors for Air Pollution Monitoring

Nowadays, most low-cost gas sensors are either metal-oxide semiconductor (MOS) gas sensors (e.g., the ozone sensor depicted in Figure 1.4) or electrochemical gas sensors (e.g., the carbon monoxide sensor depicted in Figure 1.5). These sensors are affordable (essential in participatory sensing scenarios) and due to their form factor highly suited for mobile measurements. Below, we briefly describe the functional principle of both gas sensor types and discuss their limitations.

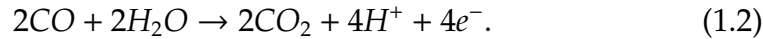
Semiconductor gas sensors. Semiconductor gas sensors are electrical conductivity sensors. The impedance characteristic of the active sensing layer is altered through the contact with the gas to be detected. A wide range of gases can be monitored with semiconductor sensors, such as surface ozone (O_3), carbon monoxide (CO), and nitrogen oxides (NO_x , which is the sum of nitric oxide NO and nitrogen dioxide NO_2). In the ideal case, the oxidation and reduction processes at the sensing layer are totally reversible. For example, the sensing mechanism of ozone can be modeled as



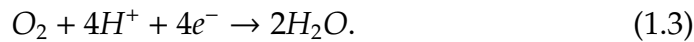
where e^- is a conduction electron from the sensitive layer and O^- is a surface oxygen ion. The result is a reduced electron flow in the sensitive layer resulting in the increase of its electric resistance. This resistance of the sensitive layer is measured to derive the ozone concentration.

Electrochemical gas sensors. Electrochemical gas sensors consist of two electrodes (a sensing electrode and a counter electrode), which are separated by an electrolyte. The target gas reacts at the sensing

electrode involving an oxidation or reduction process. This generates a current between the two electrodes, which is proportional to the gas concentration. For example, when measuring carbon monoxide the reaction at the sensing electrode can be modeled as



At the same time there is a reaction at the counter electrode



Hence, the overall reaction is $2CO + O_2 \rightarrow 2CO_2$, depicting that a small amount of oxygen (O_2) is required to complete the reaction. For this purpose electrochemical gas sensors have a vent hole on the back of the sensor cell.

Limitations. The main drawbacks of both types of low-cost gas sensors are their limited accuracy and resolution, low stability, and poor selectivity. Low-cost gas sensors are usually installed in industrial production factories to measure high-level concentrations (e.g., in the automotive industry [PPS⁺99]). However, the monitoring of environmental gases in ambient air requires to measure very low gas concentrations. The supported ranges by the gas sensors are often 1000x larger than the typical concentration ranges of the pollutants in ambient air, which amplify the problem of limited accuracy and resolution in the intended monitoring range.

Additionally, low-cost gas sensors are unstable [HZG11]. Due to sensor aging, they must be re-calibrated every month [DVMP⁺08, TYIM05] or even every week [KBP06]. Another great challenge is the poor selectivity of the sensors. The sensor output is sensitive to ambient humidity and temperature, and responsive to influences of interfering gases [KBP06]. Frequent sensor calibration can compensate for part of these effects in order to improve or at least preserve the desired measurement accuracy [BMEP03, BN07, MLCOS08, LB14].

Regularly calibrating sensors, analyzing the quality of the measurements collected, and deriving thereof useful information for others (e.g., environmental scientists, epidemiologists, and general public) are challenging tasks. In the following, we discuss the arising difficulties.

1.3.3 Post-Processing of the Measurements

As discussed above, low-cost gas sensors need periodic calibrations to retain a high measurement accuracy. There are two common approaches for the calibration of gas sensors. The calibration can take place in the laboratory using artificial gas mixtures [CKCH09] or in the field

with real pollution measurements by placing the sensor close to a fixed station providing reliable measurements [BM98, CMC⁺01, KBP06]. The disadvantage of calibrating the sensors with real pollution measurements is the dependency on weather conditions and local pollution dispersion, which both cannot be controlled. For a precise calibration, reference measurements under a wide range of environmental conditions are desirable, e.g., low, middle, and high concentrations of the target pollutant under distinct humidity and temperature settings [TYIM05]. Nevertheless, usually, calibrating sensors in the field is the only feasible option, since the sensors are part of a measurement systems (e.g., installed in smartphones) and cannot be taken into a laboratory every time a calibration is required. Hence, automatic calibration algorithms are needed to keep calibration parameters up-to-date. This is particularly challenging in networks with uncontrolled mobility, where sensors can not be forced to pass by certain locations, e.g., to stay close to a fixed station to use its reliable measurements for sensor calibration.

Low-cost sensors are increasingly used in applications scenarios where wrong or inaccurate sensing may lead to wrong decisions with crucial societal and economical impacts [CCD⁺11, LLL⁺09, CLBR10]. For example, on days with high air pollution levels the government could decide to restrict speed limits on highways, impacting the mobility of people and goods. Therefore, for many applications it is relevant to not only assess the absolute value of the sensor readings, but also gain information about the quality of individual measurements. This is a challenging task as the measurement quality depends on many factors, such as type of sensor used, sensor aging effects, calibration quality, and environmental conditions. In addition, quality estimates for individual sensor readings enable the detection of measurement outliers and help to track down defective sensors.

1.3.4 Providing Useful Applications

Data users are in general not interested in single pollution measurements but rather in a consistent view of the pollution distribution in a certain region. Hence, we need methods to accurately model the distribution of urban air pollutants at a high temporal and spatial resolution. Node mobility trades off temporal resolution against spatial resolution, enabling a high spatial resolution across large areas with a small number of sensor nodes. However, due to the reduced temporal resolution of any covered location, it is a formidable challenge to derive pollution maps with a high temporal resolution (e.g., with a daily or hourly resolution).

The public availability of reliable pollution maps is essential. They raise the citizens' awareness, empower environmental scientists to craft

and evaluate new policies, and enable in combination with other spatial data sets (e.g., road network, population density, and real estate prices) a wide range of novel applications. Different modeling approaches exist to derive pollution maps [JAK⁺04]. Simple interpolation-based methods only require the concentration and location of the measurements while complex physical dispersion models need a rich set of supplementary data, such as detailed geometric building models. It is important to choose the right model depending on the required accuracy, resolution, and available input data.

1.4 Thesis Contributions

The four main contributions of this thesis are:

- We are the first to study multi-hop calibration in detail. We propose a *novel multi-hop calibration algorithm*, which exploits co-located sensor measurements to enhance sensor calibration. Our algorithm *highly reduces error propagation* in the network and considerably outperform existing multi-hop calibration algorithms.
- We introduce models to describe the phenomenon of interest and the sensor hardware and propose a model-driven approach to compute *accuracy bounds* for noisy sensor readings. Additionally, we can automatically *detect* systematic and transient *sensor errors*.
- We integrate gas sensors in mobile sensor nodes, which are deployed on top of public transport vehicles to collect spatially highly resolved data sets of urban air pollutants. We use the collected data to develop a *novel modeling approach*, which enables us to derive *pollution maps* with an *unprecedented spatio-temporal resolution*.
- The derived high-resolution pollution maps enable new types of applications and provide valuable information to environmental scientists, epidemiologists, and the general public. We integrate the derived pollution maps in a smartphone application, which *helps urban dwellers to assess and reduce* their exposure to air pollutants.

In the following, we detail the outline of the thesis and separately list the contributions of every chapter.

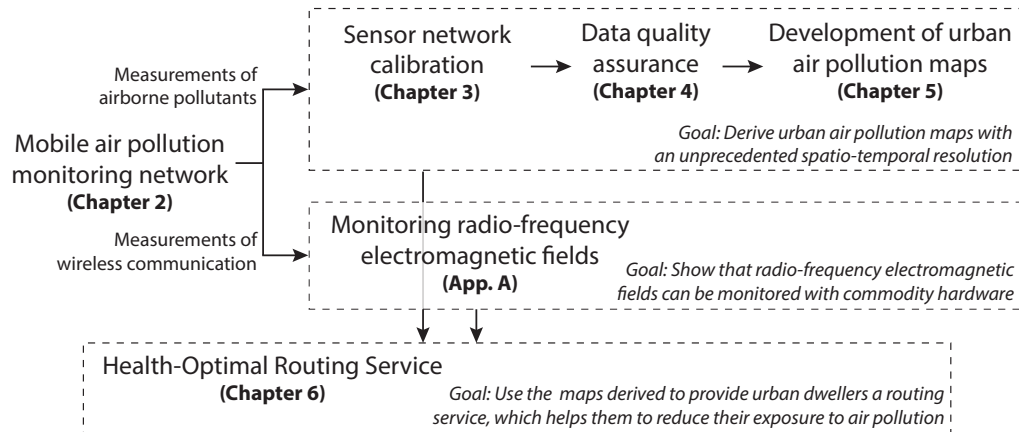


Figure 1.6: Organization of this thesis.

1.5 Thesis Outline

The organization of this thesis is depicted in Figure 1.6. We present the hardware systems and software algorithms required to accurately monitor and reason about phenomena, which are spatially highly variable and, hence, can not be profoundly monitored by fixed measurement stations only. In particular, the main scope of this thesis is to increase the knowledge about the spatio-temporal distribution of air pollutants in urban areas. We introduce a mobile air pollution monitoring network (Chapter 2) to monitor urban air pollutants with a high spatial resolution. We use the collected measurements to build pollution maps with an unprecedented spatio-temporal resolution (Chapters 3–5). Finally, we use the derived pollution maps to provide urban dwellers a health-optimal routing service, which helps them to reduce their exposure to air pollution (Chapter 6). Besides analyzing common air pollutants, we also explore whether useful inferences can be drawn from spatially resolved monitoring of electromagnetic pollution on the GSM and Wi-Fi frequency bands (App. A).

The contributions of the individual chapters are as follows:

Chapter 2: Mobile Air Pollution Monitoring Network

This chapter introduces the mobile air pollution monitoring network developed to monitor a wide range of pollutants with a high spatial and temporal resolution. We present the design of two independent measurement platforms, which together constitute a powerful and flexible air pollution monitoring network.

- We describe the architecture, implementation, and evaluation of a city-scale mobile measurement system based on ten sensor nodes deployed on top of streetcars of the local public transport network.

- We introduce a low-power mobile sensing system for participatory air quality monitoring. The prototypical implementation interfaces a low-cost gas sensor with an off-the-shelf Android smartphone.

Chapter 3: Calibrating a Mobile Network of Low-Cost Sensors

This chapter tackles the main drawbacks of low-cost gas sensors, which are limited sensibility, low stability, and poor selectivity. We propose a novel multi-hop calibration algorithm, which is able to calibrate large, heterogeneous, mobile sensor networks. Hence, we can achieve a good measurement accuracy despite the limitations of the low-cost gas sensors deployed in our sensor platforms.

- We introduce a novel network-wide calibration algorithm with small multi-hop error propagation influence.
- We investigate the accuracy of our calibration algorithm with a detailed theoretical analysis, simulations under various settings, and real pollution measurements from our mobile sensor network.

Chapter 4: Accuracy Bounds for Low-Cost Sensor Readings

This chapter presents a model-driven approach to provide accuracy bounds for calibrated measurements from low-cost sensors and to detect systematic and transient sensor errors.

- We introduce generic models to describe a phenomenon of interest and the characteristics of the sensor in use. We present an algorithm, which uses these models to compute reliable accuracy bounds for individual sensor readings and detect erroneous measurements.
- We demonstrate the feasibility of the method proposed by analyzing large sets of calibrated sensor readings from static and mobile sensor networks.

Chapter 5: Developing Fine-Grained Air Pollution Maps

This chapter makes use of calibrated and cleaned measurements from the mobile monitoring network to derive urban air pollution maps with a high spatial and temporal resolution.

- We derive land-use regression models with a spatial resolution of 100 m x 100 m and use standard metrics to analyze the quality of the models for different temporal resolutions.
- We propose a novel modeling approach to increase the quality of pollution maps with high (sub-weekly) temporal resolutions.

Chapter 6: The Health-Optimal Route Planner

This chapter makes use of the generated high-resolution pollution maps to provide urban dwellers a health-optimal routing service. Citizens can reduce their exposure to air pollution by not taking the shortest path between origin and destination but a healthier and slightly longer alternative route.

- We analyze the occurring trade-off between the lower exposure to air pollution, which is payed by a longer travel distance.
- We develop a smartphone application, which computes shortest and health-optimal paths between two arbitrary locations in the city.

Besides common air pollutants, the deployed sensor nodes also monitor various wireless communication frequency bands.

Appendix A: Monitoring Radio-Frequency Electromagnetic Fields

Radio-frequency electromagnetic fields are emitted by applications using wireless communication. A part of the general public is increasingly concerned about the long-term effects of electromagnetic radiation on human health. This chapter examines whether expensive state-of-the-art exposure meters can be replaced with commodity sensor nodes to simplify the monitoring of radio-frequency electromagnetic fields.

- We demonstrate that radio-frequency electromagnetic field measurements are feasible with off-the-shelf smartphones and the commodity sensor nodes of our mobile measurement network.
- We show that the measurement accuracy is similar to that of state-of-the-art exposure meters used in medical studies.
- We use the data collected to develop a fine-grained GSM exposure map with a high spatial resolution of 100 m x 100m.

2

Mobile Air Pollution Monitoring Network

In recent years, wireless sensor networks (WSNs) have become a mature technology and are successfully used in a number of long-term installations [BGH⁺09, CMP⁺09]. Most deployments are composed of static sensor nodes fixed at desired locations, which are carefully chosen according to the node's communication and sensing ranges. The data gathered by such deployments is usually highly temporally resolved at every covered location. The main drawback of static installations is, however, their poor spatial scalability to monitor large areas, e.g., an entire city.

In many urban areas, public transport vehicles, such as streetcars and buses, provide a suitable basis for deploying wireless sensor nodes [BM98, ASC⁺10, DSL⁺13, RPV14]. The inherent mobility and coverage of public transport systems enables monitoring a large area with a few mobile sensor nodes. We monitor airborne pollutants and environmental parameters with sensor nodes installed on top of streetcars of the public transport network in the city of Zurich, Switzerland.

The measurements gathered by fixed stations and vehicle-based sensor networks may not adequately represent the exposure situation of individual persons. We tackle this challenge by additionally introducing a community-driven sensing infrastructure. Given the broad availability of personal GPS-equipped smartphones, we aim to use these devices to build a large-scale sensor network of mobile devices for participatory air pollution monitoring [SD10].

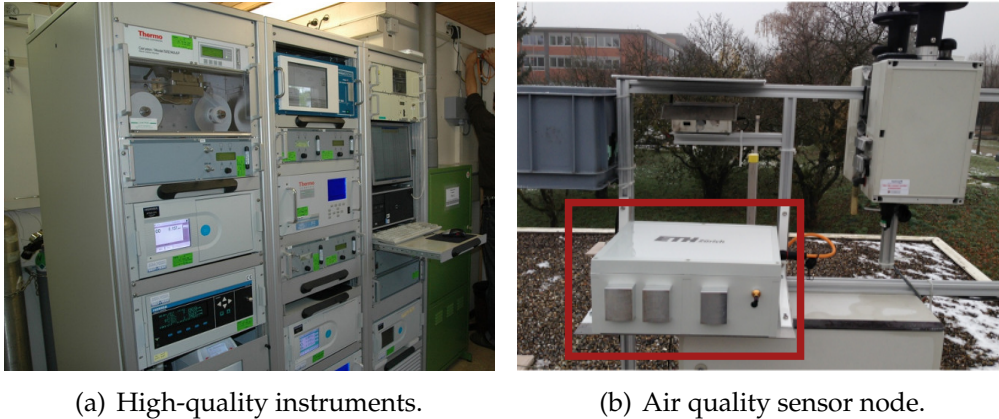


Figure 2.1: The air quality sensor node is installed on top of a fixed station (a) of the NABEL network close to the station’s air inlets (b). Every sensor reading from the sensor node can be compared to a high-quality reference measurement.

Contributions. This chapter has the following contributions. First, we introduce our long-term sensor testbed, which we use to evaluate the quality of low-cost gas sensors. Only sensors, which perform well are installed in our sensor nodes. Second, we describe our system design consisting of two parts:

- A compact air quality sensor node, which is able to monitor the concentration of different air pollutants (e.g., ultrafine particles (UFP), carbon monoxide (CO), and nitrogen dioxide (NO₂)) along with temperature and humidity conditions. This sensor node is designed to be deployed on top of mobile vehicles.
- A small and portable prototypical participatory measurement device, composed of an off-the-shelf smartphone and a low-cost gas sensor to monitor the ozone (O₃) concentration.

Finally, we describe the different deployments where our sensor nodes are in operation and detail the data sets collected.

2.1 Long-Term Sensor Testbed

Low-cost gas sensors are either delivered uncalibrated or with a factory calibration. The factory calibration is often not intended for low concentration measurements as found when monitoring urban air pollutants. Usually, the calibration parameters are only based on measurements of a small number of different gas concentrations under one specific temperature and humidity setting [Alp14a, Alp14b].

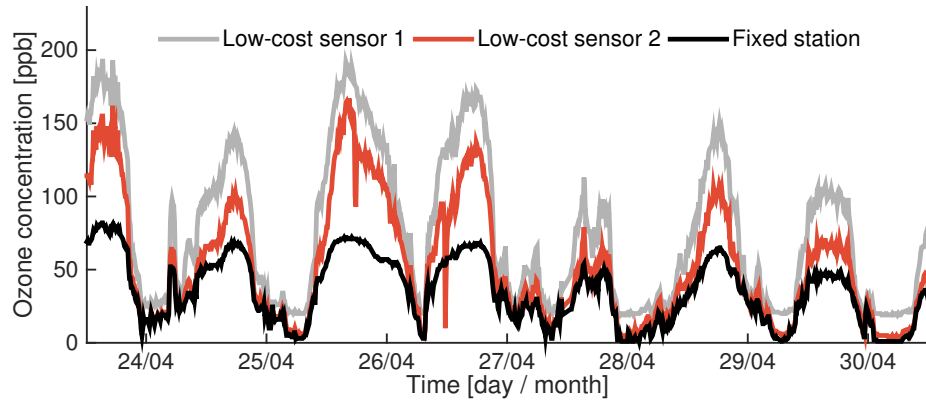


Figure 2.2: Measurements from two factory calibrated semiconductor ozone sensors compared to reference measurements from the fixed station.

Possibly, the chosen setting is not suitable for the intended measurement campaign (e.g., different conditions apply for indoor and outdoor monitoring of pollutants). Additionally, low-cost gas sensors have limited accuracy and resolution, low stability, and poor selectivity, i.e., the sensors' output is sensitive to ambient humidity and responsive to influence of interfering gases [KBP06]. Hence, it is essential to assess the quality of the sensor readings with tests in the field, i.e., with real urban pollution measurements.

For this purpose we operate since April 2011 a sensor testbed infrastructure. We installed a reconfigurable and easily extendable sensor node with multiple gas sensors on top of a fixed station of the National Air Pollution Monitoring Network (NABEL), as depicted in Figure 2.1. We use this side-by-side operation of low-cost sensors and high-quality instruments as long-term sensor testbed to analyze the performance of low-cost gas sensors. For example, Figure 2.2 shows the measured ozone concentration from two factory calibrated ozone gas sensors and the fixed station over a period of one week. We see that the measurements are off the mark by up to a factor of two. However, the sensor readings show a very high correlation with the reference signal (Pearson correlation of 0.96 and 0.97, respectively). This indicates that using more suitable calibration parameters will highly improve the accuracy of the measurements. Only sensors which perform well with respect to measurement accuracy, reliability, and response time, are finally installed in the sensor nodes of the mobile measurement network, as we will describe in Section 2.2.

The sensor testbed also allows us to evaluate the long-term signal stability of the deployed sensors. For example, one low-cost sensor is continuously monitoring the ozone concentration since more than three years by now, giving useful insights about the sensor's performance under a variety of different conditions.

Data set	Measurements [in millions]	Sampling interval	Time period
Ozone (O ₃)	1.7	60 s	3.5 years (ongoing)
Temperature / Humidity	1.7	60 s	3.5 years (ongoing)
Carbon monoxide (CO)	2.4	10 s	9 months (ongoing)
Nitrogen dioxide (NO ₂)	2.4	10 s	9 months (ongoing)

Table 2.1: Data sets collected by the static sensor node installed on top of a fixed measurement station (from April 2011 to August 2014).

2.1.1 Data Sets

Table 2.1 shows the different data sets collected with the sensor testbed over the past 3.5 years (from April 2011 to August 2014). We tested many different sensors over this time span, but we only list measurement from sensors that are able to deliver a good measurement quality and, hence, are installed in the sensor nodes of the mobile sensor network. For every measurement listed in the table, we have a corresponding high-quality reference signal from the fixed station.

Within our sensor testbed, gas sensors are tested in a static setting. However, the gas sensors finally selected are deployed in a mobile sensor network. Therefore, we also analyze how mobility, specifically the wind speed, affects sensor readings.

2.1.2 Effect of Mobility on Sensor Readings

In the following, we analyze how sensor mobility effects the measurements of semiconductor gas sensors (we did not observe any influence for electrochemical gas sensors). Semiconductor gas sensors have a sensing layer, whereof the resistance is measured to derive the concentration of the monitored pollutant. This resistance R is heavily temperature dependent. Hence, each sensor reading includes not only the value of the resistance R but also the on-board temperature T to compute the temperature-compensated resistance \tilde{R}

$$\tilde{R} = R \cdot e^{K(T-T_0)}, \quad (2.1)$$

with reference temperature $T_0 = 25^\circ\text{C}$ and temperature coefficient K , e.g., for the ozone sensor the coefficient is $K = 0.025$ [SGX14].

We analyze the effect of sensor mobility on the accuracy of the sensor readings, which is due to the varying air flow around the sensor head. We carry out several experiments in a closed room with constant ozone concentration. We use a table fan, which generates a maximum wind speed of 6.6 m/s to analyze the influence of the air flow on the raw sensor

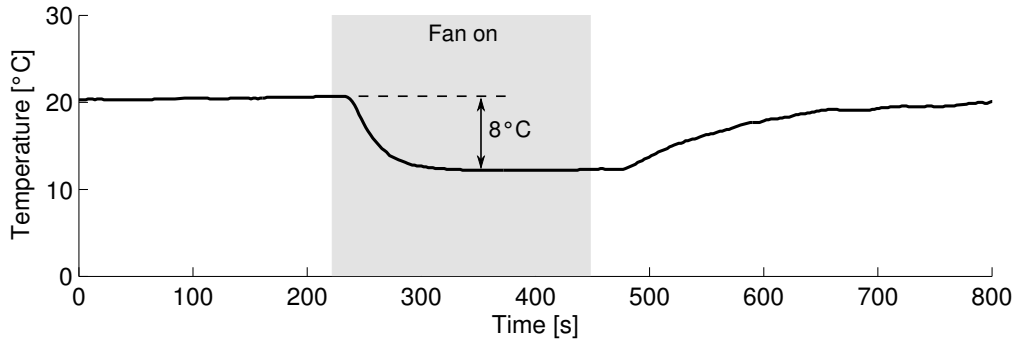


Figure 2.3: Air flow generated by a fan (shaded area) influences the readings of the on-board temperature sensor. A maximum drop of 8°C is measured.

readings. We observe that the air flow mainly impacts the on-board temperature T , which is used to calculate the temperature-compensated resistance \tilde{R} of the sensing layer. The air flow around the sensor head influences the heat dissipation on the sensor board and results in a lower temperature reading of at most $T_a = 8^{\circ}\text{C}$, as shown in Figure 2.3. Using the temperature coefficient of the ozone sensor, we derive that the temperature drop induces a maximum relative error of 14% in the calculation of the temperature-compensated resistance:

$$1 - \tilde{R}_a/\tilde{R} = 1 - e^{-K \cdot T_a} = 0.14. \quad (2.2)$$

This maximum relative difference is negligible for low ozone concentrations, but results in a high offset under high pollution levels. No precaution is required for measurement campaigns with pedestrians, which are usually moving at slow speeds. However, we recommend to protect the sensor head from a direct exposure to air flow under rapid motion speeds of the sensor head, e.g., on top of mobile vehicles and while riding the bicycle. Alternatively, accelerometer data can be used to measure motion speeds in order to compensate the temperature drop due to mobility.

2.2 System Architecture

In the following, we describe our system architecture for mobile air pollution monitoring. First, we depict the design of our vehicle-mounted sensor node, which can be used to automatically monitor large urban areas. Second, we introduce our low-power mobile sensing subsystem

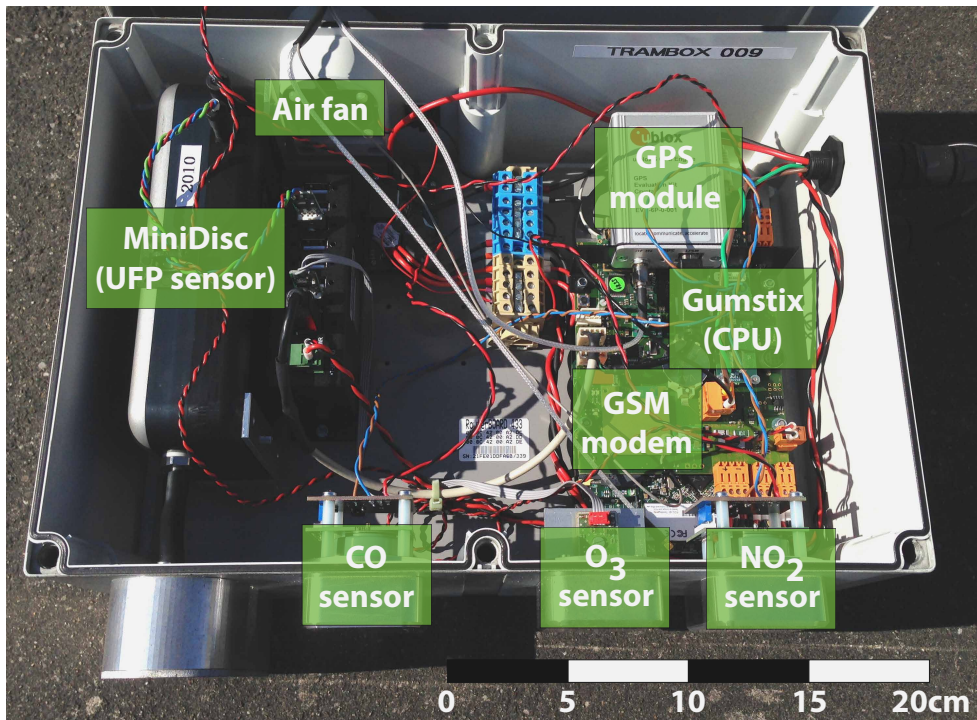
for participatory air pollution monitoring. Finally, we briefly outline the back-end infrastructure.

2.2.1 Vehicle-Mounted Air Quality Sensor Node

The air quality sensor node is based on the sensor platform [BYL⁺11] developed within the PermaSense [BGH⁺09] and X-Sense [BBF⁺11] projects for high-alpine permafrost monitoring. We adapted and extended this platform to enable mobile air pollution monitoring. The core of the sensor node, depicted in Figure 2.4, is a Gumstix embedded computer with a 600 MHz CPU running the Ångström embedded Linux operating system [Gum14]. A GPS receiver supplies the station with precise geospatial information [Ub14]. The station supports bidirectional communication over GSM and WiFi on the 2.4 GHz and 5 GHz frequency bands [Gem14, Mik14]. GSM is used for data transmission under regular system operation while WiFi is used during debugging and maintenance phases. All radio signals (GPS, GSM, and WiFi) are processed by a single planar antenna, which is mounted on top of the box as shown in Figure 2.4(b). The weight of a node is approximately 4.5 kg and it has a maximum power draw of around 40 W. The sensor node requires an external power source, e.g., from the vehicle on which it is mounted on.

All air intakes are equipped with protection covers, as depicted in Figure 2.5. The bottom side of the aluminum cover has multiple air vents and two inclined bars to retain water and dust. Additionally, the cover protects the sensor heads from a strong direct air flow, which would influence the sensor readings, as discussed in Section 2.1.2. The fan on the back of the node draws air out of the box to ensure a steady air flow. To monitor air pollution, the sensor nodes are equipped with a semiconductor ozone (O₃) sensor [SGX14], electrochemical carbon monoxide (CO) and nitrogen dioxide (NO₂) sensors [Alp14a, Alp14b], and a novel compact device to measure ultrafine particle (UFP) concentrations [FHSB11]. Additionally, the nodes monitor radio-frequency electromagnetic fields using the GSM modem and the WiFi access point (more details in App. A) and environmental parameters, such as temperature and humidity [Sen14].

We use Python scripts to communicate with the different sensors over their serial communication port. All measurements are annotated with time and location information and then stored on the local memory card in an SQLite database. The data packets are transmitted to the back-end server as soon as a connection is available (either through GSM or WiFi). The sensor readings are removed from the local database once their reception is acknowledged by the back-end server.



(a) Inner view.



(b) External view.

Figure 2.4: The air quality sensor node is equipped with UFP, CO, O₃, and NO₂ sensors. Geotagged and time-stamped measurements are transmitted over GSM (cellular network) to the back-end server for further processing.

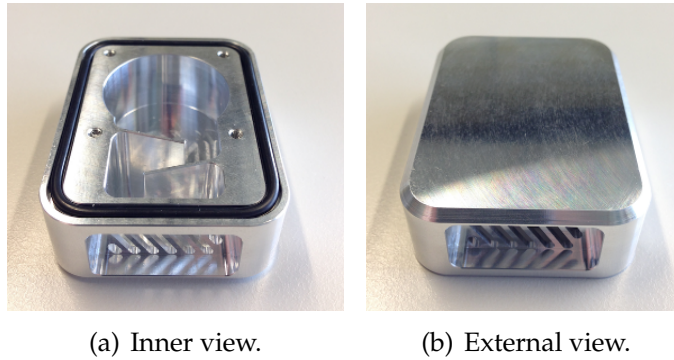


Figure 2.5: All air intakes have protection covers to retain water and dust.

2.2.2 Participatory Sensing Using Smartphones

We introduce our prototypical air quality measurement system GasMobile, which interfaces a small-sized, low-cost ozone sensor with an off-the-shelf smartphone running the Android OS. In the following, we describe the hardware and software system designs.

Hardware architecture. Our measurement system consists of four parts as displayed in Figure 2.7(a). We use a semiconductor-based ozone sensor from SGX Sensortech [SGX14] (same as the one in the air quality sensor node) to sense the ozone concentration in the atmosphere based on the measured resistance of the sensor’s tin dioxide (SnO_2) layer. Digital communication is possible over the board’s RS232-TTL interface, which is directly connected to an HTC Hero smartphone providing a USB Mini-B port. All parts are off-the-shelf hardware available for low prices (in the range of hundreds of dollars in total). This is essential to obtain widespread acceptance of participatory sensing equipment.

Having an extended battery lifetime is crucial for mobile and participatory sensing applications. We analyze the total current draw of the ozone sensor and the USB-RS232 translator, both components being powered by an external battery pack.¹ We use an Agilent digital multimeter with a sampling rate of 100 ms; the measured current draws are illustrated in Figure 2.6. After each power-on, the tin dioxide layer of the ozone sensor is overheated for 60 s. This overheating decreases the sensor drift over time. The current draw during the overheating phase is 47 mA. After overheating, the sensor is ready for taking measurements. We put the sensor in automatic mode in which it uses its own clock to automatically perform measurements every two seconds. This ensures that always an up-to-date measurement reading is available whenever the smartphone application decides to request a sensor reading. Each

¹Today’s smartphones support USB host mode and, hence, can power peripherals.

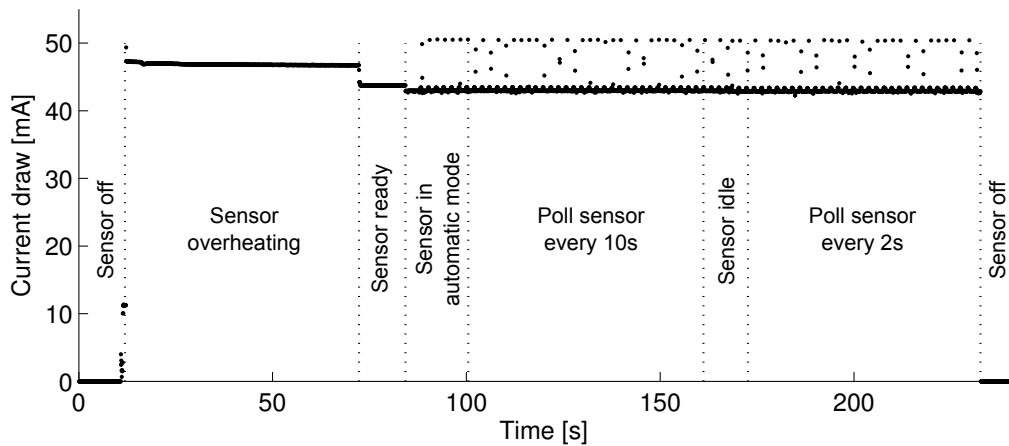
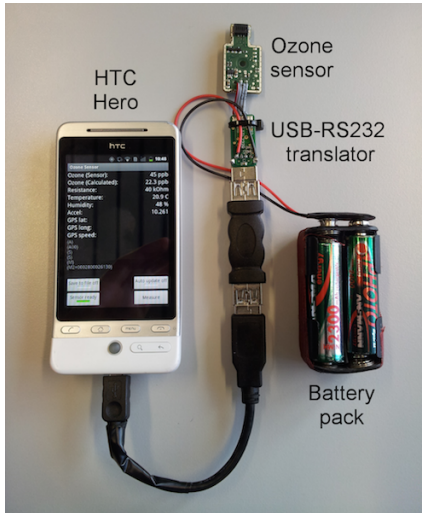


Figure 2.6: Current draw of the ozone sensor and USB translator over time. Sensor polling does not noticeably increase the current draw.

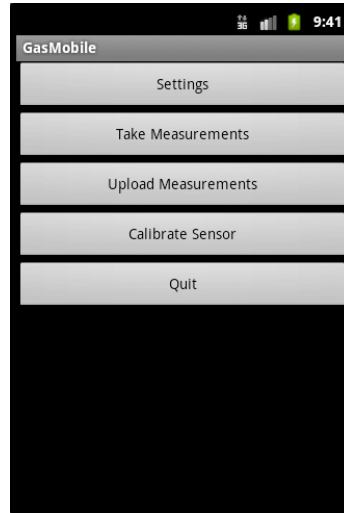
measurement results in a short 50 mA peak of the current draw, as shown in Figure 2.6. Applications polling sensor readings do not noticeably increase the current draw.

We operate the gas sensor using four AAA NiMH batteries with a nominal capacity of 2500 mAh at 1.2 V. Considering the highest measured current draw of 50 mA (240 mW), we roughly estimate a battery lifetime of 50 hours. This lifetime allows us to monitor the ozone concentration for approximately one month, assuming that an adult spends on average 1.7 hours per day outdoors [FBG⁺02]. Most standard operations on today’s smartphones use more energy, e.g., sending an SMS amounts to 302 mW, a phone call over GSM needs 1054 mW [CH10].

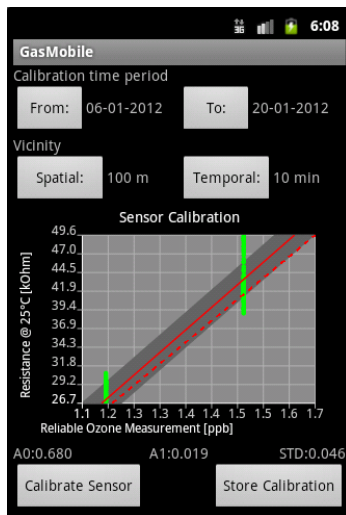
Android application. The application starts with the main menu depicted in Figure 2.7(b). The user can access the settings, take measurements, calibrate the sensor, or upload the measurements to a server. In the settings screen, the user can change several configuration parameters. In the measurements screen (see Figure 2.7(d)) the user can put the sensor in automatic mode and choose whether to poll the sensor once or continuously with a pre-configured poll interval. The application polls the latest raw data from the ozone sensor (resistance, temperature, and humidity), and position and speed information from the GPS module. The ozone concentration is calculated and displayed in the plot on the screen. The geo-localized and time-stamped measurements can be permanently stored on the smartphone’s memory card and uploaded to the back-end server for further processing and visualization.



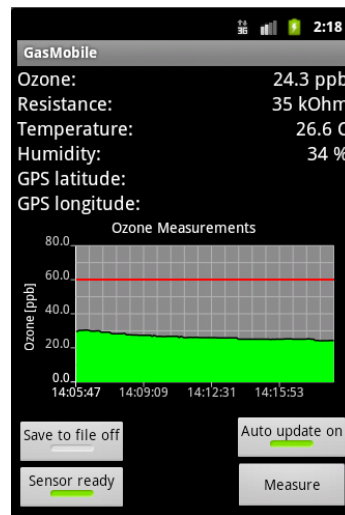
(a) Hardware architecture.



(b) Main menu.



(c) Calibration.



(d) Measurements.

Figure 2.7: GasMobile hardware architecture (a) and Android application (b)-(c). The user can set the poll interval, adjust calibration parameters, poll sensor measurements, and upload the measurements to a server for further processing.

Extensibility to other gas sensors. Extending GasMobile to support other sensors is straightforward and only requires minor modifications in two software components, as long as the sensor provides serial communication over USB. First, the serial communication protocol has to be tailored to the software and hardware requirements of the intended sensor. Second, the Android application must be implemented to facilitate the interaction between user and sensor.

2.2.3 Back-end Infrastructure

The back-end infrastructure is shared with the PermaSense and X-Sense projects [ETH14]. We use the same setup but adapt the data processing units to handle air pollution data. On the back-end server we run two instances (private and public) of the Global Sensor Network (GSN) [AHS06] software package, a Java-based software middleware that facilitates data collection in sensor networks. The sensor nodes communicate with the private GSN instance. It is used to store all the measurements obtained in a MySQL database without any further processing of the data. This enables us to re-process the data if needed and serves as backup of the original raw measurements received from the sensor nodes.

The private GSN instance streams the received data to the public GSN instance where we post-process the measurements, e.g., by aggregating different data streams, filtering out outliers, and updating calibration coefficients. The post-processed data are stored in a separate MySQL database. The public GSN instance also offers a publicly accessible webpage, which allows researchers and the general public to download the measured pollution concentrations and enables us to monitor the proper functioning of the sensor nodes and to visualize the spatial distribution of the monitored air pollutants.² We use Vizzly [KBST12] for the need to handle increasingly larger data volumes. Vizzly is a web-based caching and visualization tool, which allows us to quickly display and analyze large data sets on top of Google Maps.

2.3 Deployments and Collected Data Sets

This section introduces three deployments where the air quality sensor node and smartphone device described above are in operation.

2.3.1 Deployment 1: Streetcar Measurement Network

The main deployment of this thesis is formed by a network of air quality sensor nodes installed on top of vehicles of the public transport network in Zurich, Switzerland. Starting in the year 2012, we gradually equipped ten streetcars with our sensor nodes, as shown in Figure 2.8(a). The nodes are mounted on top of a platform to allow an optimal air flow to the sensitive layer of the gas sensors. The instrumented streetcars cover on a regular schedule a large urban area of 100 km².

²<http://data.opensense.ethz.ch>

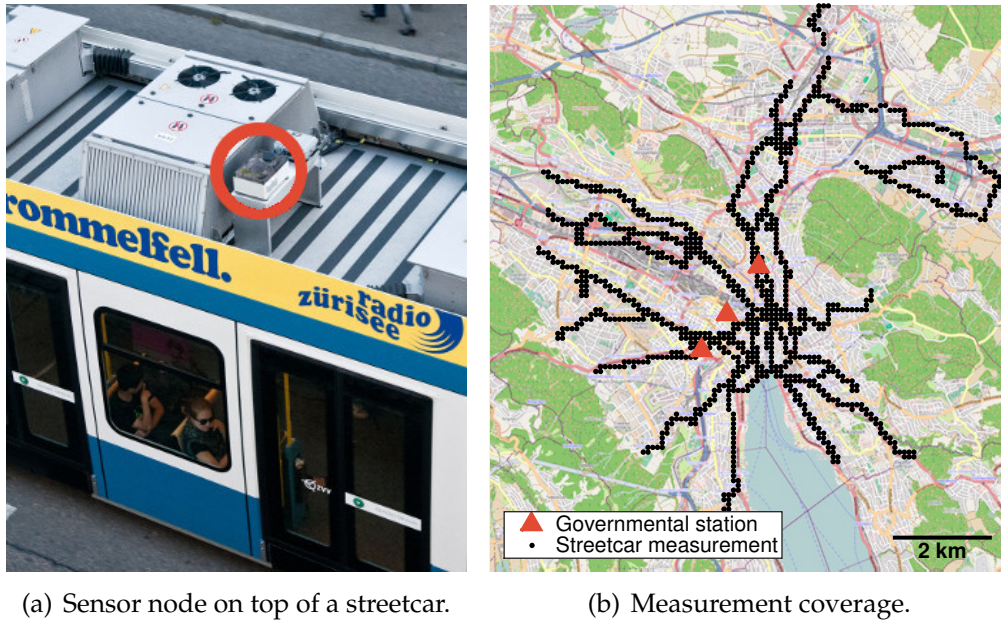


Figure 2.8: Ten air quality sensor nodes are installed on top of streetcars of the public transport network in Zurich, Switzerland. The sensor nodes achieve a good coverage within an urban area of 100 km².

The spatial coverage of the sensor nodes' measurements is shown in Figure 2.8(b). The dots denote locations (100 m x 100 m) with at least 50 measurements over the course of the last 2.5 years (from April 2012 to August 2014). The ten mobile sensor nodes achieve a good coverage of the region, in particular compared to the three governmental stations denoted with triangles in Figure 2.8(b). One station belongs to the NABEL network and two stations are part of the cantonal measurement network OSTLUFT. These two stations are located very close to the streetcar tracks with a distance of 4 m and 16 m, respectively. Hence, we can use their high-quality measurements to adjust the calibration parameters and assess the measurement accuracy of the passing sensor nodes (for more details see Chapter 3 and Chapter 4).

We show in Figure 2.9(a) the operation times of the individual nodes. A sensor node is on average 20 hours per day in operation. During the night, typically from 1 AM to 5 AM, the streetcars are in their depots and the nodes are powered off. The streetcars equipped with our sensor nodes collect air pollution data on 80% of the days on average, as depicted in Figure 2.9(b). As it can be seen in the figure, not all sensor nodes were deployed at the same time. We had two installation phases, the first five sensor nodes were mounted in April 2012 and the remaining five ten months later in February 2013. A wide range of different factors contribute to the difference between total number of days deployed and

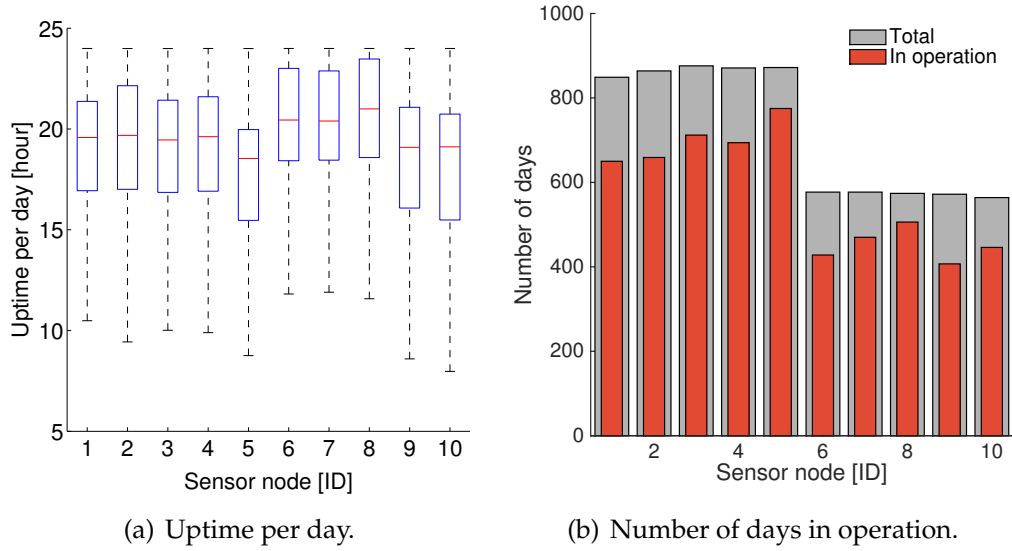


Figure 2.9: The sensor nodes on top of the streetcars are on average 20 hours a day and 80 % of the days in operation.

Data set	Measurements [in millions]	Sampling interval	Time period
Ultrafine particle (UFP)	69	5 s	2.5 years (ongoing)
Ozone (O ₃)	11	20 s–60 s	2.5 years (ongoing)
Temperature / Humidity	11	20 s–60 s	2.5 years (ongoing)
Carbon monoxide (CO)	9	10 s	4 months (ongoing)
Nitrogen dioxide (NO ₂)	9	10 s	4 months (ongoing)
RF electromagnetic fields	4	15 s–30 s	4 months (finished)

Table 2.2: The ten mobile sensor nodes collected a wide range of pollution measurements over the course of two years.

number of days in operation. Streetcars may be out of service for several reasons: (i) The public transport company has more streetcars than the number of vehicles needed on a regular day, (ii) streetcars have regular maintenance phases and may be out of operation for several weeks after an accident, and (iii) also the air quality sensor nodes contribute to the downtime by having maintenance, repair, and upgrade phases.

We collected a diverse set of spatially highly resolved pollution measurements, as summarized in Table 2.2. We have large data sets of ultrafine particle (UFP), ground ozone (O₃), carbon monoxide (CO), and nitrogen dioxide (NO₂) measurements. Additionally, the nodes monitor radio-frequency electromagnetic fields and environmental parameters, such as temperature and humidity. In total, we collected more than 110 million pollution concentration measurements annotated with accurate timing and location information from all over the city of Zurich.

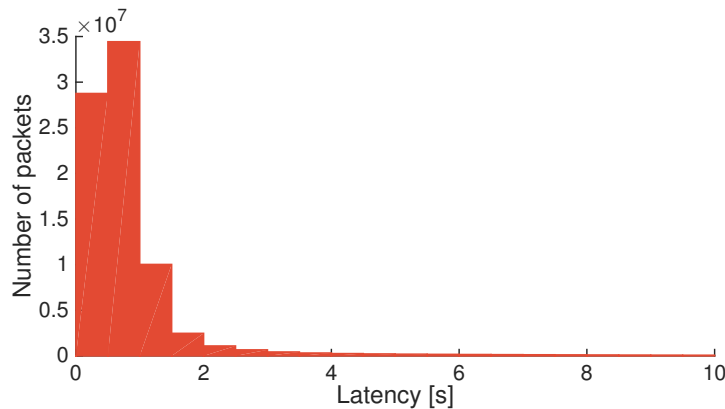


Figure 2.10: The latency of most data packets of the streetcar deployment is below 2 s, which makes real-time services feasible.

We finish the description of this deployment by showing the data latency, defined as the time passed between data generation on the sensor node and data reception at the back-end server. Low data latency is key to enabling real-time applications, e.g., to provide a timely warning system for days with high exposure levels. The sensor nodes on top of the streetcars are connected to the GSM network most of the time and, hence, can immediately transmit the monitored concentrations to the back-end server. The median latency is at 0.8 s. The maximum latency is as high as multiple weeks due to back-end server upgrades leading to offline times of several weeks. Figure 2.10 shows the latency distribution for all packets with a latency below 10 s. It can be seen that most of the data packets have a latency below 2 s.

2.3.2 Deployment 2: LuftiBus Measurements

Our second mobile deployment is in cooperation with Lunge Zurich [Lun14], which is a health organization for lung diseases such as asthma, sleep apnea, and tuberculosis. Lunge Zurich consults and treats patients and cooperates with universities for clinical trials. Additionally, Lunge Zurich strives to reduce the number of patients with lung diseases by supporting and realizing prevention and health check campaigns.

We cooperate within the LuftiBus campaign. LuftiBus (depicted in Figure 2.11(a)) is a bus equipped with instruments for measuring the functioning of lungs to detect possible dysfunctions. The goal is to prevent and early diagnose lung-related diseases. The LuftiBus can be rented by schools, universities, cities, companies, and events (e.g., sports meetings). People interested can then use the medical equipment and the available professional staff of the LuftiBus to test their lung capacities and get

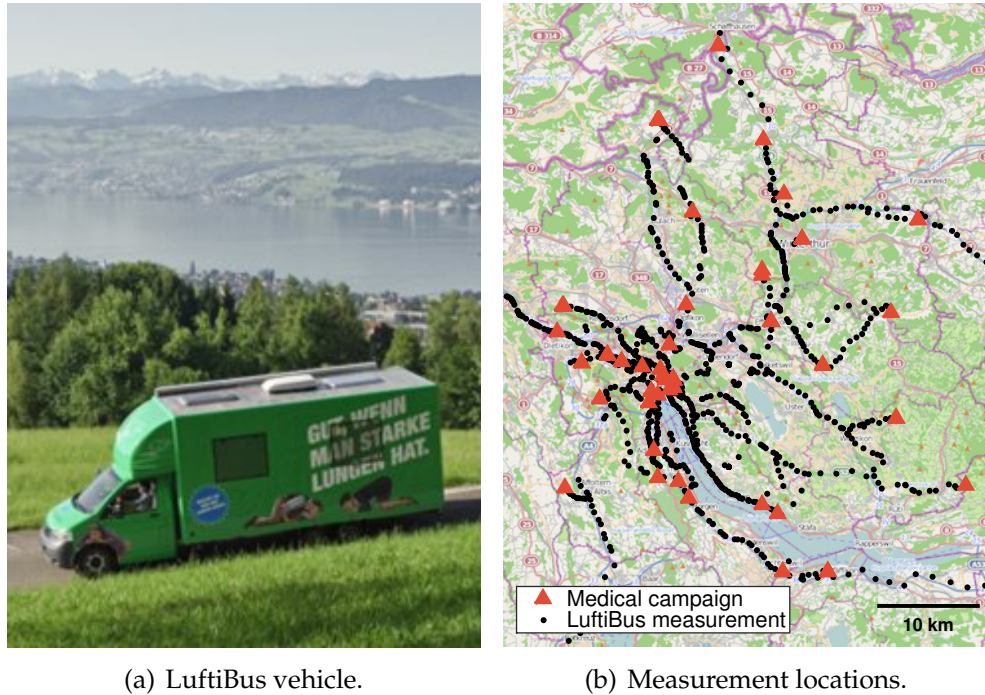


Figure 2.11: The air quality sensor node is installed on top of the LuftiBus vehicle making measurements all around Zurich (displayed in (b)) and Switzerland. Long measurements periods were possible at the medical campaign sites.

feedback of their measurement results.

We installed one of our air quality sensor nodes on top of the LuftiBus vehicle. This allows us to monitor the air quality while the bus is on the road and at locations it is stopping for the medical campaigns. The aim is to analyze the correlation between the medical test results and the measured air pollution concentration in the region. We collect air pollution measurements on top of LuftiBus since January 2013. The LuftiBus visited during the last 1.5 years over 50 locations, mainly around the city of Zurich as depicted in Figure 2.11, but also at some other locations in Switzerland. The triangles show locations with conducted

Data set	Measurements [in millions]	Sampling interval	Time period
Ultrafine particle (UFP)	1.3	5 s	1.5 years (ongoing)
Ozone (O ₃)	0.12	60 s	1.5 years (ongoing)
Temperature / Humidity	0.12	60 s	1.5 years (ongoing)
Carbon monoxide (CO)	0.13	10 s	6 months (ongoing)
Nitrogen dioxide (NO ₂)	0.13	10 s	6 months (ongoing)

Table 2.3: Data sets collected on top of the LuftiBus vehicle.

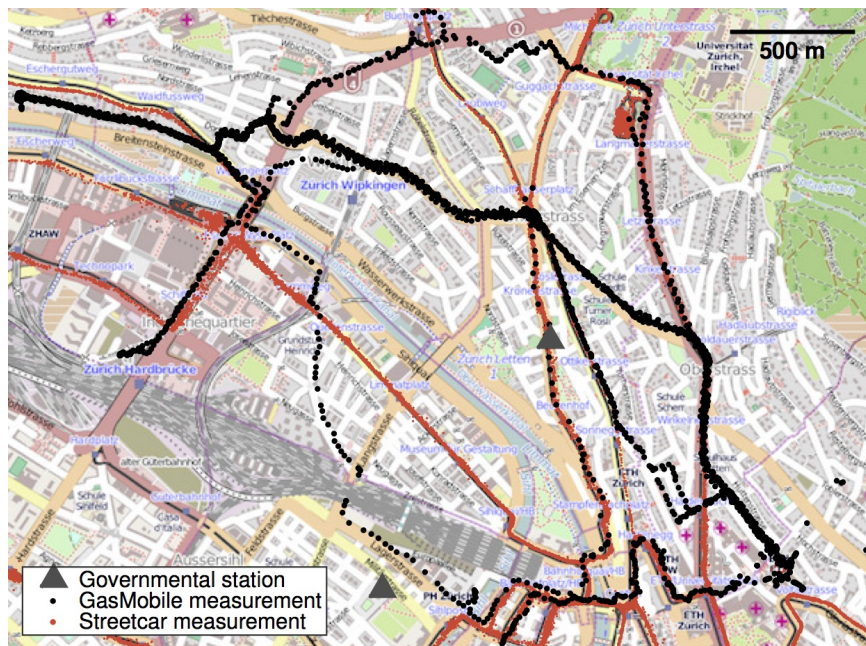


Figure 2.12: GasMobile pollution measurements from several bicycle rides with a polling interval of five seconds. The quality of the measurements can be assessed by comparing them to data collected by governmental stations and the streetcar network.

medical campaigns where our sensor node was able to monitor air pollutants over longer time periods (from multiple hours up to several days). Table 2.3 lists the amount of data collected for every monitored pollutant. So far, we collected around 2 million pollution measurements.

2.3.3 Deployment 3: Participatory Pollution Monitoring

We have available a full system for mobile participatory sensing, ranging from the sensing hardware and client software as described in the previous sections to a powerful web-based visualization tool to analyze the measured concentrations on top of Google Maps. In the following, we briefly present results from a measurement campaign using GasMobile.

We used GasMobile during multiple days to monitor the ozone concentration in the city of Zurich, Switzerland. For this, we mounted the smartphone with the connected sensor on a bicycle (protected from wind) and took measurements from several bicycle rides at different locations in the city, as shown in Figure 2.12. Throughout the measurement campaign we used a sampling interval of five seconds and collected in total 2,800 spatially distributed data points. All sensor readings were directly uploaded to the back-end server running GSN. The location- and time-based caching tool Vizzly enables users to efficiently retrieve the

data and allows combining the measurements with data sets gathered by other sensor networks. Measurements collected in the vicinity of governmental stations and the streetcar network (see Figure 2.12) can be used to update GasMobile’s calibration parameters and to assess its data quality, as described in Chapter 3 and Chapter 4.

2.4 Related Work

Vehicle-based sensing. Mobile vehicles are very useful to collect a wide range of spatially resolved measurements enabling novel types of applications.

Taxicab networks are well-suited to collect spatially resolved sensory data in metropolitan areas. Often large number of taxicabs operate within a dense, highly populated area (e.g., in New York City over 13,000 taxicabs provide around 485,000 trips per day [Tax14]) enabling a fine-grained coverage of the region. Additionally, taxicabs often automatically upload vehicle status and location information to central dispatching centers to optimize routes for passenger pick ups. The data collected from taxicab networks can be used for a wide range of applications. Aslam et al. [ALPR12] infer traffic volumes through data collected from a sensor network of taxicabs. The authors propose models and inference procedures to detect traffic patterns in real-time. Chen et al. [CZZ⁺13] and Shang et al. [SZT⁺14] use taxicab GPS traces to optimally plan routes for a night bus network and to infer gas consumption and pollution emission of vehicles, respectively.

Furthermore, also other means of transport are used to collect sensory data. Völgyesi et al. [VNKL08] use car-mounted sensor nodes to collect air pollution data. They introduce a prototype platform to monitor ozone (O₃), carbon monoxide (CO), and nitrogen dioxide (NO₂). The measurements are periodically uploaded to a server and displayed as contour maps. Mathur et al. [MJK⁺10] equip cars with ultrasonic range-finder sensors and GPS receivers to determine parking spot occupancy. The sensed data is aggregated at a central server to create a real-time map of parking availability. Kapoor et al. [KHLH14] make use of airplanes as large-scale sensor network for identifying wind speeds on a continental scale. The authors assess air and ground speeds reported by aircrafts at different locations and altitudes to infer the prevailing wind speeds and directions.

We are not the first with the vision to use public transport vehicles for air pollution monitoring. Devarakonda et al. [DSL⁺13] propose two data farming models to collect air pollution data. One that can be deployed

on public transportation and a second that can be used as a personal sensing device. The authors present preliminary prototypes and discuss implementation challenges. Re et al. [RPV14] envision to extend the network of fixed air quality stations in Palermo, Italy, with mobile bus-mounted sensor nodes for air quality monitoring.

In comparison to the works described above, we are the first to demonstrate that it is feasible to reliably operate a mobile air quality sensor network for multiple years collecting millions of useful measurements throughout a large urban area.

Participatory sensing. There is a plethora of new application scenarios, which are enabled by participatory sensing networks. Thiagarajan et al. [TBGE10] introduce a cooperative transit tracking application using data collected by the public transport user's smartphones. The goal is to improve rider experience by reducing waiting times at bus stations. Faulkner et al. [FOC⁺11] make use of acceleration data collected with smartphones and community-owned sensors to detect emerging earthquakes allowing to issue real-time alerts. Rana et al. [RCK⁺10] and Meurisch et al. [MPSS13] use smartphone sound recordings to collect noise pollution data to create noise maps, while Zheng et al. [ZLW⁺14] use data from a public platform where citizens can report incidences (such as loud music) to create a noise map of New York City. Ganti et al. [GPA⁺10] use participatory sensing data of measurements of vehicular fuel consumption sensors to map fuel consumptions on city streets. Koukoumidis et al. [KPM11] use smartphones to detect and predict traffic signal schedules to provide drivers and their onboard computational devices with information about the schedule of the traffic signals ahead. This enables the drivers to optimally adjust their speed or take a detour if required to minimize fuel consumption.

Smartphones are also widely used to facilitate data collection. For example, Miluzzo et al. [MLF⁺08] use smartphones for visibility monitoring, Mohan et al. [MPR08] to monitor road and traffic conditions, Rachuri et al. [RMM⁺10] for sensing individual emotions, and Lai et al. [TTLLSC11] for bicycle localization.

Many of these smartphone-based sensing applications either do not need any external sensors or use bluetooth for data transfer between sensor and smartphone [FKS⁺11, JLT⁺11, HBPW08, DAK⁺09]. Bluetooth gives the user great freedom in sensor placement, but leads to higher battery drain due to bluetooth communication on the device and sensor side. We instead exploit USB host mode and directly connect the sensor to the smartphone and therewith reduce the power consumption of the measurement system.

Monitoring air pollution using low-cost gas sensors has gained high

interest in recent years [TYIM05, ZBP⁺12, KEH⁺13, JJWF13]. Low-cost gas sensors are often embedded in custom-build sensor nodes that are part of mobile sensor networks [HBPW08, CKCH09, DAK⁺09]. Instead, we control the gas sensor with minimal additional hardware using an off-the-shelf smartphone. This keeps material costs low and thus makes our measurement system attractive to a large number of people as a large-scale sensor network of mobile phones [KGZ07].

2.5 Conclusions

In this chapter, we introduced our mobile air quality measurement network consisting of a continuously running vehicle-based sensor network and a prototypical smartphone device suitable for participatory sensing scenarios.

Our mobile air quality sensor nodes can monitor a wide range of highly relevant air pollutants. The sensor nodes are in operation at three different deployment sites. Our main installations consists of ten air quality sensor nodes deployed on top of ten streetcars of the public transport network in Zurich, Switzerland. The instrumented vehicles are traversing a large urban area on a regular schedule collecting large data sets of spatially resolved air pollution data. From April 2012 to August 2014, we collected more than 110 million time-stamped and geo-localized pollution measurements. With the LuftiBus deployment, we facilitate the investigation of a possible correlation between lung function and air quality and with our smartphone-based GasMobile prototype system, we show that participatory air pollution monitoring is feasible.

In the following, we examine the quality of the gathered measurements and propose strategies and algorithms, which enable the calibration of a network of low-cost gas sensors (Chapter 3) and allow assessing the quality of calibrated gas sensor measurements (Chapter 4).

3

Calibrating a Mobile Network of Low-Cost Sensors

Frequent sensor calibration is essential in sensor networks with low-cost sensors. We exploit that temporally and spatially close measurements of different sensors measuring the same phenomenon are similar. Hence, when calibrating a sensor, we adjust its calibration parameters to minimize the differences between co-located measurements of previously calibrated sensors. A freshly calibrated sensor can now be used to calibrate other sensors in the network, referred to as multi-hop calibration.

We are the first to study multi-hop calibration with respect to a reference signal (micro-calibration) in detail. We show that ordinary least squares regression—commonly used to calibrate noisy sensors—suffers from significant error accumulation over multiple hops. In this chapter, we propose a novel multi-hop calibration algorithm using geometric mean regression, which *(i)* highly reduces error propagation in the network, *(ii)* distinctly outperforms ordinary least squares in the multi-hop scenario, and *(iii)* requires considerably fewer ground truth measurements compared to existing network calibration algorithms. The proposed algorithm is especially valuable when calibrating large networks of heterogeneous sensors with different noise characteristics. We provide theoretical justifications for our claims. Then, we conduct a detailed analysis with artificial data to study calibration accuracy under various settings and to identify different error sources. Finally, we use our algorithm to accurately calibrate large data sets of air pollution measurements gathered by our mobile air pollution monitoring network.

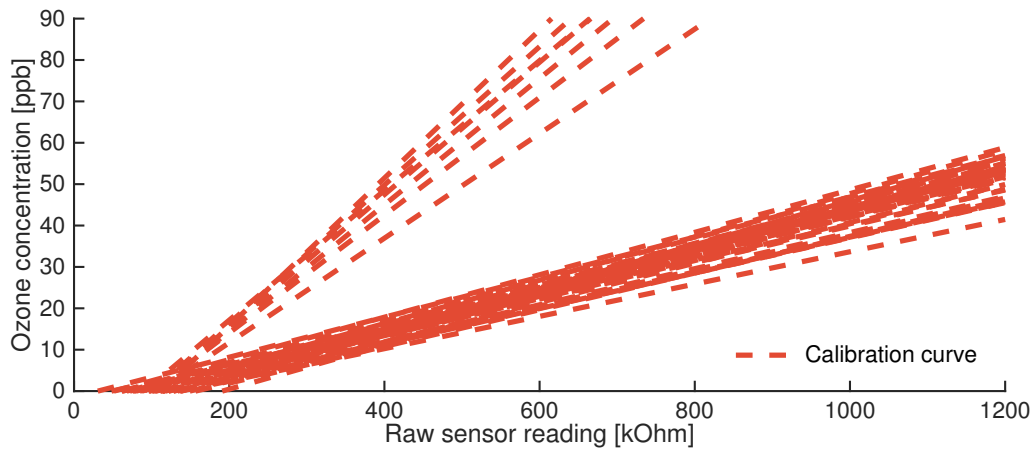


Figure 3.1: Calibration curves of one low-cost ozone sensor over the period of one year. For every week of the year the optimal calibration curve is depicted.

3.1 Introduction

Wireless sensor networks (WSNs) are increasingly used in a wide range of application domains to gather information about the physical world. Nowadays, WSNs are embedded in real-world deployments that go beyond research prototypes. Examples include the monitoring of permafrost in high-alpine regions [BGH⁺09], surveillance of a heritage building in Northern Italy [CMP⁺09], and air pollution monitoring in urban regions [LFS⁺12]. All these installations have been in operation for multiple years. Such long-term deployments need to function correctly over long time periods without requiring frequent maintenance phases. In addition, in many of these deployments the quality of the gathered data is vital for the success of the whole application. Data collected by WSNs are used for adaptive lighting in road tunnels [CCD⁺11], data center monitoring [LLL⁺09], and clinical patient surveillance [CLBR10]. In these application scenarios wrong or inaccurate sensing may lead to wrong decisions with significant societal and economic impact.

Challenges. Achieving a good data quality and preserving it during the whole system lifetime is essential. However, in most application scenarios this is a very challenging task. Deployment specifics and limited budgets often constrain the choice of sensing hardware. Typical limits apply to sensor size, price, and energy consumption. Advances in sensor technology constantly widen the spectrum of phenomena that can be captured with WSNs by bringing small, cheap, and portable sensors onto the market. The downside of this trend is a reduced accuracy, precision, and reliability of many available sensors [MSPD09]. One common example are sensors produced for air quality monitoring. In recent years, low-cost gas sensors (~100 dollars) appeared on the market, which

are compact in size and suitable for mobile air pollution measurements. However, compared to traditional instruments, many gas sensors have a very limited accuracy, e.g., [SGX14, Alp14a, Alp14b]. Furthermore, many low-cost sensors suffer from sensor drift, have limited stability, and are sensitive to changing environmental conditions [RBB⁺06, HZG11]. Figure 3.1 shows a typical example how the sensitivity of low-cost sensors fluctuates over time. The figure depicts the calibration curves of one low-cost ozone sensor over the course of one year. We derive the sensor’s optimal calibration curve for every week of the year, using reference measurements from our sensor testbed described in Section 2.1. There is no clear indication where the bipartition of the calibration curves is coming from. Hence, it is evident that low-cost sensors need to be frequently calibrated to preserve a good data quality [MLCOS08, BN07, LB14]. Manual calibration is an elaborate and time-consuming task [TPS⁺05, RBB⁺06]. Automatic sensor calibration is essential but challenging, particularly in networks with uncontrolled mobility.

Problem statement. In the context of this work, a fundamental observation is that two measurements taken in the spatial and temporal vicinity of each other are similar. The required spatio-temporal closeness between two measurements depends on the locality of the process monitored. If two sensors fulfill the vicinity requirements, we term their meeting point a *rendezvous*. Rendezvous between calibrated and uncalibrated sensors can be used to calibrate uncalibrated sensors. These sensors in turn, can calibrate other uncalibrated sensors. We refer to this as *multi-hop sensor calibration*.

One-hop calibration, i.e., the calibration between a high-quality reference signal and an uncalibrated sensor, is common practice. State-of-the-art network calibration algorithms designed for mobile sensing networks only perform one-hop calibration [MLCOS08, MLE⁺15]. Hence, these algorithms can only calibrate sensors, which directly pass by a high-quality reference sensor. This introduces a strong constraint, requiring a high density of reference sensors, which is not feasible for many application scenarios. Moreover, many large-scale monitoring systems have limited and uncontrollable mobility patterns preventing sensors to frequently pass by a reference sensor, e.g., sensors installed in private smartphones [MPSS13].

The only work we know of that studies multi-hop calibration by leveraging meeting points between uncalibrated sensors is [BMEP03]. Further, in our previous work we also analyze multi-hop calibration, where our initial simulation results reveal a linear error accumulation with increasing network size [HST12]. Both works use ordinary least squares regression (*OLS*) to calibrate a network of sensors. *OLS* is

typically used to map uncalibrated raw sensor readings to calibrated measurements if the dependency between target signal and sensor response is linear [LUWP11, BN07, MLE⁺15]. In this chapter, we show that, despite the linear dependency between target signal and sensor response, *OLS* is not appropriate for multi-hop calibration. We propose a novel multi-hop calibration algorithm, which distinctly outperforms existing works by considerably reducing error propagation in the network.

Contributions and road-map. We are the first to study multi-hop calibration with respect to a reference signal in detail. This work presents a thorough theoretical analysis based on assumptions and models introduced in Section 3.2 and Section 3.3. We analyze the sources of error propagation and propose a network calibration scheme, which (i) considerably reduces error propagation in the multi-hop setting, (ii) distinctly outperforms multi-hop calibration with *OLS*, and (iii) compared to existing work does not require high density of reference sensors. The algorithm proposed leverages geometric mean regression (*GMR*) and one of its main properties: absence of regression dilution bias, caused by noise in sensor readings [Woo41].

In summary, this chapter makes the following contributions:

- We analyze in Section 3.4 the *OLS* line fitting method—commonly used to calibrate sensors—and introduce *GMR* line fitting, the centerpiece of our multi-hop calibration algorithm.
- In Section 3.5, we present a detailed theoretical analysis of calibration error accumulation in a network of uncalibrated sensors. We prove that multi-hop calibration with *GMR* is optimal under specific assumptions. Further, we show with simulations that *GMR* outperforms *OLS* in many scenarios.
- We show in Section 3.6 how to best choose calibration paths in dense networks, where uncalibrated sensors can calibrate themselves to possibly multiple calibrated sensors.
- We run in Section 3.7 the proposed network calibration algorithm on real data of more than 13 million air pollution measurements from our streetcar deployment. We show the benefits of calibrating real, noisy, and low-cost sensors with our novel algorithm.

The algorithm proposed enables the multi-hop calibration of large, heterogeneous, and mobile sensor networks, e.g., as found in participatory sensing scenarios [BEH⁺06]. Until now, this has not been possible with existing network calibration algorithms. We survey related work in Section 3.8, and conclude in Section 3.9.

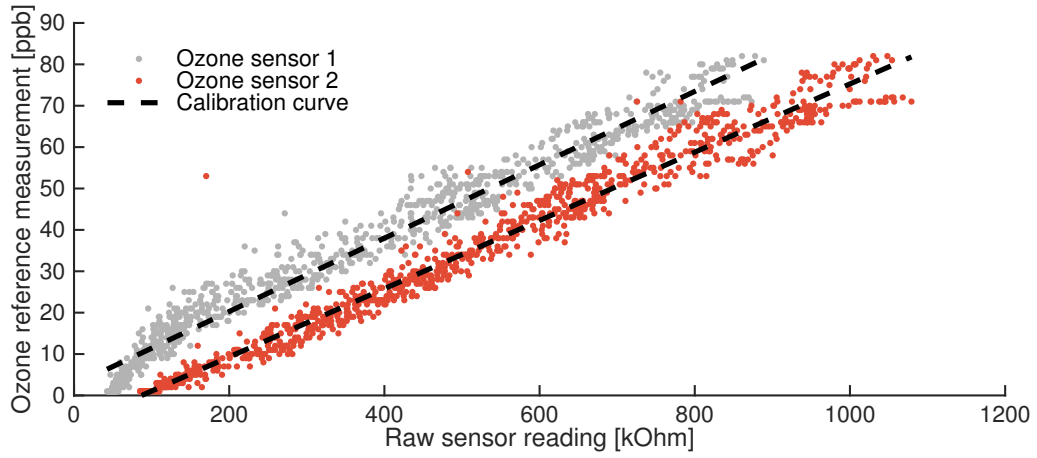


Figure 3.2: Two MiCS-OZ-47 ozone sensors with a linear dependence between raw sensor readings and reference measurements.

3.2 Assumptions and Models

Let a monitoring system consist of a set of sensor nodes (or sensors) U measuring a phenomenon H . A sensor node can be either static or mobile. Each node knows its position and current time. A sensor $u \in U$ takes a time-ordered sequence of measurements and meets other sensors from time to time. There are no constraints on the sensor sampling rate and on the frequency of rendezvous between any two sensors.

3.2.1 Phenomenon Model

Let a phenomenon of interest H exhibit a continuous measurable signal $h: T \times L \rightarrow D$ at time $t \in T \subseteq \mathbb{R}^+$, at location $l \in L \subseteq \mathbb{R} \times \mathbb{R}$ defined by latitude and longitude coordinates, and domain of measurable values $D \subseteq \mathbb{R}$. We assume that the process change is upper bounded in both time and space, i.e., for any two locations $(t_i, l_i), (t_j, l_j) \in T \times L$ in time and in space, it holds that

$$|h(t_i, l_i) - h(t_j, l_j)| \leq f_\omega(|t_i - t_j|, |l_i - l_j|), \quad (3.1)$$

where f_ω is a *monotonically non-decreasing* function of temporal and spatial distances $|t_i - t_j|$ and $|l_i - l_j|$. The slower f_ω grows and the smaller the distances $|t_i - t_j|$ and $|l_i - l_j|$ are, the more *similar* are the values of h at the two time-space locations (t_i, l_i) and (t_j, l_j) .

3.2.2 Sensor Model

A sensor $u \in U$ takes a sequence of measurements $\{m_u(t_i, l_i)\} \in D_u$ of a phenomenon of interest H from the domain of values D_u possibly aperiodically at discrete time-space locations $(t_i, l_i) \in T \times L$, $i \in \mathbb{N}$. We consider a measurement as point measurement, that is, it has no duration. A sensor u is *perfect* if at any point in time and space $(t, l) \in T \times L$ it holds that $m_u(t, l) = h(t, l)$. We call h the *phenomenon signal* and denote with m_u the raw measurements of sensor u .

Many low-cost sensors show a close to linear dependence between sensor values m_u and phenomenon signal h , e.g., [SGX14, Alp14a, Alp14b]. For example, we depict in Figure 3.2 one week of measurements from two semiconductor ozone sensors [SGX14] gathered with our sensor testbed infrastructure. Both sensors show a linear dependence between high-quality reference measurements and raw sensor readings. We assume for all sensors $\forall u \in U$ that

$$h = \alpha_u + \beta_u m_u + e_u, \quad (3.2)$$

where α_u and β_u are *calibration coefficients* of the first order polynomial and e_u is the *noise* component (or *sensor error*) with zero mean [BMEP03, BN07]. This noise component includes all variations not explained by the calibration curve [Smi09]. We assume that the errors of any two sensors are independent. The calibration parameters α_u and β_u describe intercept (offset) and slope (gain) of the calibration curve. We calibrate a sensor $u \in U$ by assigning it calibration parameters. A calibrated measurement is then

$$\widehat{m}_u = \alpha_u + \beta_u m_u. \quad (3.3)$$

We distinguish two types of sensors: *reference sensors* $u^* \in U^*$, which are calibrated at all times, and *low-cost sensors* $v \in U \setminus U^*$, which require periodic calibration. We require (infrequent) rendezvous between at least one reference sensor and one low-cost sensor to calibrate a network of low-cost sensors.

3.3 Rendezvous Between Sensors

We perform sensor calibration by leveraging rendezvous between sensors, defined as spatially and temporally close measurements of two sensors. This section introduces the main concepts.

3.3.1 Definition of a Rendezvous

Let $u, v \in U$ be two sensors. Both sensors take measurement sequences $\{m_u\} \subset D_u$, $\{m_v\} \subset D_v$ and eventually meet, i.e., some measurements are

in each other's vicinity. We define the set of spatially and temporally close pairs of measurements $\Phi^{(u,v)}$ between sensors u and v within a time interval s as

$$\Phi^{(u,v)} = \{ (m_u(t_i, l_i), m_v(t_j, l_j)) \mid t_i, t_j \in s \wedge |t_i - t_j| \leq \Delta t \wedge |l_i - l_j| \leq \Delta d \}. \quad (3.4)$$

The parameters Δd and Δt set temporal and spatial constraints on the required closeness of the measurements and, thus, limit the possible change of the monitored phenomenon (see (3.1)). The choice of Δd and Δt depends on the process of interest. For example, the change of ozone concentration over short distances is insignificant [SHWT13, MLE⁺15] in contrast to the concentration variability of fine particles [PTPB13]. We detail in Section 3.7 our choice of Δd and Δt for temperature, ozone (O_3), and carbon monoxide (CO) measurements.

Set $\Phi^{(u,v)}$ represents a *rendezvous* between sensors u and v . We refer to the measurement pairs $(m_u, m_v) \in \Phi^{(u,v)}$ as *calibration pairs* and use them to improve sensor calibration. We use the projection operator " \downarrow " to split the measurements belonging to calibration pairs according to the sensors by which the measurements were made. For example, the measurements of rendezvous $\Phi^{(u,v)}$ can be split into the two sets $\Phi_{\downarrow u}^{(u,v)} = \{m_u\}$ and $\Phi_{\downarrow v}^{(u,v)} = \{m_v\}$ denoting the measurements of sensor u and sensor v , respectively.

A rendezvous $\Phi^{(u,v)}$ is characterized by the parameters Δd and Δt , which control the number of calibration pairs $|\Phi^{(u,v)}|$ in the set. There is an important trade-off when choosing the values of the rendezvous parameters. On the one hand, high values of Δd and Δt result in a high number of calibration pairs $|\Phi^{(u,v)}|$, potentially giving calibration possibility to a larger set of uncalibrated sensors. On the other hand, large Δd and Δt may lead to high values of f_ω in (3.1) and, thus, to poor correlation of the measurements belonging to a rendezvous. In this case, a rendezvous $\Phi^{(u,v)}$ may include many dissimilar pairs of measurements and result in high calibration errors.

3.3.2 Rendezvous Connection Graph

A *rendezvous connection graph* Γ is an undirected graph with sensors U as set of nodes and set of edges $E = \{(u, v) \mid \forall u, v \in U, u \neq v, \Phi^{(u,v)} \neq \emptyset\}$ between nodes with co-located measurements in time interval s . We assign to each edge in E the corresponding set of measurements. We introduce set $N(u) = \{v \mid (v, u) \in E\}$ to represent the *direct neighbors* of sensor u in graph Γ , i.e., comprises sensors that have rendezvous with sensor u .

Let sensor u be calibrated and consider a rendezvous connection graph Γ with edge $(u, v) \in E$. Sensor u can calibrate sensor v (denoted as

$u \rightarrow v$) by leveraging rendezvous $\Phi^{(u,v)}$ to compute calibration coefficients α_v and β_v for sensor v . We denote the path in graph Γ used to calibrate a sensor as *calibration path*, e.g., $u \rightarrow v \rightarrow \dots \rightarrow p \rightarrow q$, and refer to sensor p , the immediate parent of sensor q , as *calibration parent* of q .

Let Γ be an arbitrary rendezvous connection graph constructed for a time interval s . A *multi-hop calibration algorithm* accepts Γ as input and assigns calibration coefficients to every uncalibrated sensor in Γ if the sensor is part of a connected component of the graph including at least one reference sensor.

Our multi-hop network calibration algorithm consists of two parts: a calibration method and a calibration parent selection strategy. Given a calibrated and an uncalibrated sensor connected with an edge in Γ , the *calibration method* defines how to compute calibration parameters for the uncalibrated sensor, which best map the uncalibrated raw sensor readings to calibrated measurements. The *parent selection* strategy defines how to best choose the calibration path between a reference sensor and an uncalibrated sensor in Γ . In the following, we present an in-depth theoretical analysis of multi-hop calibration, identify calibration error sources, and propose a calibration method and a strategy to best select calibration parents.

3.4 Calibration Methods

We start by introducing the calibration metric used throughout this chapter to (i) compare line fitting methods (Section 3.4.2), (ii) evaluate their ability to suppress multi-hop error propagation (Section 3.5), and (iii) quantify the goodness of sensor calibration (Section 3.7).

3.4.1 Calibration Metric

We use the root-mean-square error (RMSE)—a standard metric to quantify calibration errors [BMEP03, BN07]—to evaluate the calibration accuracy. We compute the RMSE between calibrated measurements \widehat{m}_v of sensor v and the corresponding phenomenon signal h as

$$RMSE_v = \left(\frac{1}{|\{\widehat{m}_v\}|} \sum (\widehat{m}_v - h)^2 \right)^{\frac{1}{2}}. \quad (3.5)$$

We know the true phenomenon signal for every generated sensor reading, when evaluating calibration methods with artificial data. For real data sets, we derive this from data collected with high-quality instruments in the temporal and spatial vicinity of \widehat{m}_v .

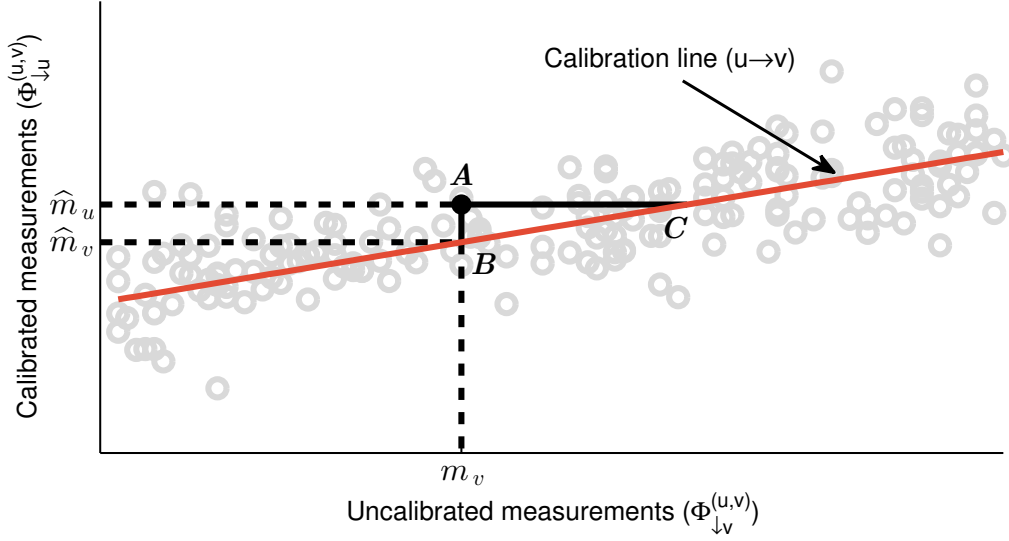


Figure 3.3: OLS minimizes the sum of squared vertical residuals AB . GMR minimizes the sum of the areas of triangles ABC formed by the vertical and horizontal residuals AB and AC .

3.4.2 Line Fitting Methods

Let sensor $u \in U$ be recently calibrated. Consider an uncalibrated sensor $v \in U$, which meets u and forms a rendezvous $\Phi^{(u,v)}$. In the following, we discuss line fitting methods that can be used to find the calibration parameters α_v and β_v of uncalibrated sensor v .

Ordinary least squares regression. Ordinary least squares (OLS) is a well-known standard method for line fitting, e.g., [BN07, LUWP11, HST12, MLE⁺15]. It is *optimal* with respect to RMSE as error metric. OLS chooses the calibration parameters α_v and β_v such that the sum of squared differences between calibrated measurements $\hat{m}_v = \alpha_v + \beta_v m_v$ of uncalibrated sensor v and measurements of calibrated sensor u is minimized:

$$\begin{aligned} & \sum_{(\hat{m}_u, m_v) \in \Phi^{(u,v)}} \left((\alpha_v + \beta_v m_v) - \hat{m}_u \right)^2 \\ &= \sum_{(\hat{m}_u, m_v) \in \Phi^{(u,v)}} \left(\hat{m}_v - \hat{m}_u \right)^2 \rightarrow \min. \end{aligned} \quad (3.6)$$

Hence, OLS minimizes the sum of squared vertical residuals, as depicted with distance AB in Figure 3.3.

OLS treats the measurements of uncalibrated sensor v and the calibrated sensor u as independent and dependent variables, respectively. Due to this distinction, two different lines regress v on u (denoted as $u \rightarrow v$)

and u on v (denoted as $v \rightarrow u$), as it makes a difference whether the residuals AB or AC are minimized in Figure 3.3. Figure 3.4 shows both regression lines $OLS(u \rightarrow v)$ and $OLS(v \rightarrow u)$ for a sample set of measurements in $\Phi^{(u,v)}$. The *true parameters* denote the used calibration parameters for sensor v to generate the artificial data. As stated in (3.6), we regress v on u to compute the slope $\beta_{OLS}^{u \rightarrow v}$ of the calibration curve (above denoted in short form as β_v):

$$\beta_{OLS}^{u \rightarrow v} = \frac{\text{cov}(\widehat{\Phi}_{\downarrow u}^{(u,v)}, \Phi_{\downarrow v}^{(u,v)})}{\text{var}(\Phi_{\downarrow v}^{(u,v)})}. \quad (3.7)$$

The slope depends on $\text{cov}(\widehat{\Phi}_{\downarrow u}^{(u,v)}, \Phi_{\downarrow v}^{(u,v)})$, which is the covariance between calibrated measurements of u and uncalibrated measurements of v at rendezvous $\Phi^{(u,v)}$ and on $\text{var}(\Phi_{\downarrow v}^{(u,v)})$, which is the variance of the latter. The covariance is independent of sensor noise e_u and e_v as they are independent variables (i.e., $\text{cov}(e_u, e_v) = 0$, $\text{cov}(\Phi_{\downarrow u}^{(u,v)}, e_v) = 0$, and $\text{cov}(\Phi_{\downarrow v}^{(u,v)}, e_u) = 0$). However, the variance in the denominator in (3.7) depends on sensor noise e_v of the uncalibrated sensor v because

$$\text{var}(\Phi_{\downarrow v}^{(u,v)}) = \text{var}(\Phi_{\downarrow v'}^{(u,v)}) + \text{var}(e_v), \quad (3.8)$$

where $\Phi_{\downarrow v'}^{(u,v)}$ denotes the noise-free measurements of sensor v . Hence, sensor noise e_v introduces a bias towards zero in the computed slope $\beta_{OLS}^{u \rightarrow v}$. The greater the noise of sensor v , the stronger the estimated slope approaches zero instead of the true slope. This is known as *regression attenuation* or *regression dilution* [FT00, CRSC06]. It is in particular a problem if sensors are calibrated over multiple hops as the bias towards zero increases with every hop. We discuss this in detail in Section 3.5.

A number of methods tackle the regression dilution problem by compensating the bias in slope estimates [Gil06, FT00]. Many of these methods assume that the variance of the sensor noise is known in advance, e.g., Deming regression [Dem44] requires that the ratio of the sensors' noise variances is known. This is particularly difficult with real sensors, which can change their characteristics over time, e.g., due to changing environmental conditions and aging effects. Using imprecise sensor noise estimates for compensating the bias may lead to worse results than not compensating at all [SP96].

Below, we present geometric mean regression, a line fitting method that does not suffer from regression dilution and does not require any knowledge on the sensor noise. It is frequently used in natural sciences (e.g., astronomy and biology) and is known under different names, such as reduced major axis [Smi09], line of organic correlation [KH50], Strömberg's impartial line [FB92], and geometric mean regression [Ric73]. The latter name is used throughout this work.

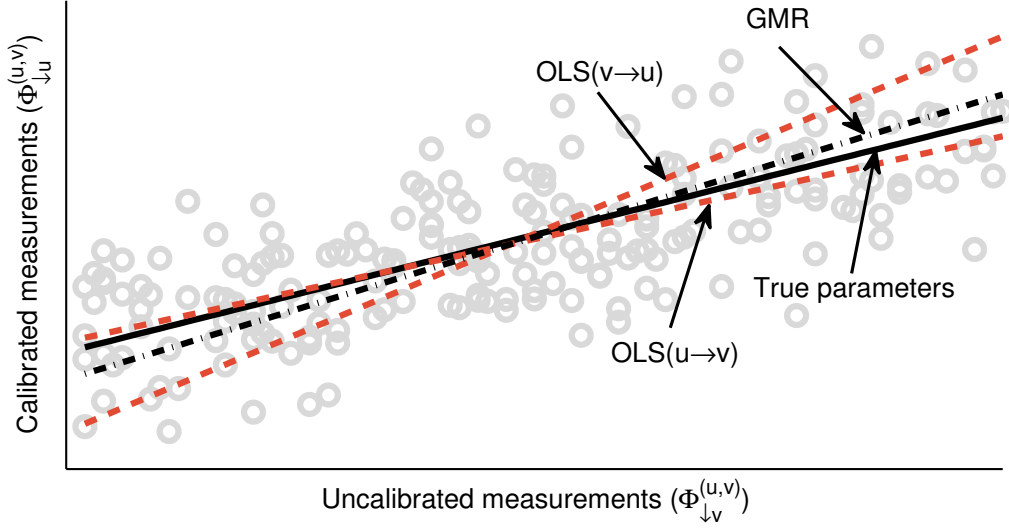


Figure 3.4: With *OLS* two different lines regress u on v and v on u . With *GMR* there is only one line as it is symmetric. The regression *OLS* ($u \rightarrow v$) is optimal with respect to the RMSE when calibrating sensor v using sensor u .

Geometric mean regression. Geometric mean regression (*GMR*) minimizes the sum of the areas of triangles formed by the deviation of a point from the regression line in both axis directions. We depict such a triangle in Figure 3.3, encapsulated by the lines AB , AC , and BC . The point A has coordinates $(\widehat{m}_u, m_v) \in \Phi^{(u,v)}$. The calibration $\widehat{m}_v = \alpha_v + \beta_v m_v$ maps raw measurements m_v to calibrated values \widehat{m}_v . Due to noise in sensor readings $\widehat{m}_v \neq \widehat{m}_u$. We can convert the calibrated values \widehat{m}_u of sensor u to the domain of raw measurements of sensor v using $-\frac{\alpha_v}{\beta_v} + \frac{1}{\beta_v} \widehat{m}_u$. Thus, the sum of areas minimized by *GMR* is

$$\begin{aligned} & \sum_{(\widehat{m}_u, m_v) \in \Phi^{(u,v)}} \left((\alpha_v + \beta_v m_v) - \widehat{m}_u \right) \cdot \left(m_v - \left(-\frac{\alpha_v}{\beta_v} + \frac{1}{\beta_v} \widehat{m}_u \right) \right) \\ &= \frac{1}{\beta_v} \sum_{(\widehat{m}_u, m_v) \in \Phi^{(u,v)}} \left(\widehat{m}_v - \widehat{m}_u \right)^2 \rightarrow \min. \end{aligned} \quad (3.9)$$

The slope $\beta_{GMR}^{u \rightarrow v}$ of the calibration curve (above denoted in short form as β_v) is

$$\beta_{GMR}^{u \rightarrow v} = \left(\frac{\beta_{OLS}^{v \rightarrow u}}{\beta_{OLS}^{u \rightarrow v}} \right)^{\frac{1}{2}} = \left(\frac{\text{var}(\widehat{\Phi}_{\downarrow u}^{(u,v)})}{\text{var}(\widehat{\Phi}_{\downarrow v}^{(u,v)})} \right)^{\frac{1}{2}}. \quad (3.10)$$

The proof of this result is given in [Woo41]. We show in Figure 3.4 the *GMR* line for a sample set of measurements. It is always located between

the regression lines $OLS(v \rightarrow u)$ and $OLS(u \rightarrow v)$, the same holds true for the line based on the true parameters.

Both OLS and GMR are least squares methods, in that they optimize the sum of squared residuals using different definitions for residuals. However, the following three properties make GMR more attractive in the context of multi-hop sensor calibration:

- *Symmetry.* Regressing $v \rightarrow u$ and $u \rightarrow v$ yields the same regression line (i.e., $\beta_{GMR}^{u \rightarrow v} = 1/\beta_{GMR}^{v \rightarrow u}$), since switching the axes does not affect the areas of triangles to be minimized. Thus, GMR does not make a distinction between dependent and independent variables.
- *Scale-invariance.* The GMR line is invariant to linear transformations of the two variables, i.e., scaling measurements of either sensor leads to an equivalent regression equation [Sam42, Kru53].
- *No bias towards zero.* In contrast to OLS , GMR does not suffer from regression dilution. This is essential to reduce calibration error propagation in the network.

Next, we investigate the error accumulation of the two line fitting methods discussed above. We investigate the worst-case scenario when noisy sensors are calibrated along a line topology with one reference sensor at one end of the line. We show that under the given assumptions GMR does not suffer from error accumulation and leads to accurate network calibration.

3.5 Reducing Error Accumulation

In the following, we provide a theoretical analysis of calibration error accumulation over multiple hops using OLS and GMR . We emphasize our findings with detailed simulation results. Our main findings are:

- OLS suffers from significant multi-hop error accumulation due to sensor noise.
- In contrast to OLS , GMR is optimal and does not suffer from any multi-hop error accumulation if the variances $var(\Phi_{\downarrow p}^{(p,q)})$, $\forall p \in U$ are accurately estimated.
- In the general case, GMR suffers from error accumulation but at a distinctly lower rate than OLS .

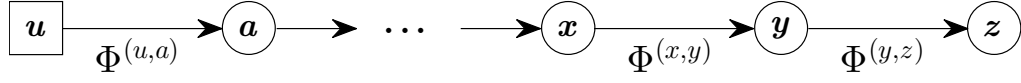


Figure 3.5: Calibration path $u \rightarrow a \rightarrow \dots \rightarrow x \rightarrow y \rightarrow z$. The sensors a, \dots, x, y, z are calibrated hop-by-hop starting from reference sensor u .

3.5.1 Theoretical Analysis

We continue with a detailed theoretical analysis.

Setting. Assume a calibration path $u \rightarrow a \rightarrow \dots \rightarrow x \rightarrow y \rightarrow z$ starting with the calibrated sensor u (reference sensor), as depicted in Figure 3.5. The line topology presented in the plot enforces sensor calibration hop-by-hop from sensor u to sensor z using the available intermediate rendezvous $\Phi^{(u,a)}, \dots, \Phi^{(x,y)}, \Phi^{(y,z)}$. All sensors but u and z have two direct neighbors. For example, sensor y has two neighbors $N(y) = \{x, z\}$, and the projections $\Phi_{\downarrow y}^{(x,y)}$ and $\Phi_{\downarrow y}^{(y,z)}$ are sets of measurements of sensor y taken in the vicinity of its direct neighbors x and z , respectively. In the general case, we can not control sensor mobility patterns and can not choose when sensors meet. Thus, the measurements in both sets have different means and variances. We use this setup to analyze multi-hop error accumulation of the previously discussed line fitting methods.

OLS multi-hop error accumulation. We analyze the calibration error of sensor z , which receives its calibration over multiple hops starting from reference sensor u . We derive the slope of the calibration curve $\beta_{OLS}^{u \rightarrow z}$ using (3.2) and (3.7) as

$$\begin{aligned}
 \beta_{OLS}^{u \rightarrow z} &= \frac{\text{cov}(\widehat{\Phi}_{\downarrow y}^{(y,z)}, \Phi_{\downarrow z}^{(y,z)})}{\text{var}(\Phi_{\downarrow z}^{(y,z)})} \\
 &= \frac{\text{cov}(\alpha_{OLS}^{u \rightarrow y} + \beta_{OLS}^{u \rightarrow y} \cdot \Phi_{\downarrow y}^{(y,z)}, \Phi_{\downarrow z}^{(y,z)})}{\text{var}(\Phi_{\downarrow z}^{(y,z)})} \\
 &= \beta_{OLS}^{u \rightarrow y} \cdot \frac{\text{cov}(\Phi_{\downarrow y}^{(y,z)}, \Phi_{\downarrow z}^{(y,z)})}{\text{var}(\Phi_{\downarrow z}^{(y,z)})} \\
 &= \dots \\
 &= \frac{\text{cov}(\Phi_{\downarrow u}^{(u,a)}, \Phi_{\downarrow a}^{(u,a)})}{\text{var}(\Phi_{\downarrow a}^{(u,a)})} \dots \frac{\text{cov}(\Phi_{\downarrow x}^{(x,y)}, \Phi_{\downarrow y}^{(x,y)})}{\text{var}(\Phi_{\downarrow y}^{(x,y)})} \cdot \frac{\text{cov}(\Phi_{\downarrow y}^{(y,z)}, \Phi_{\downarrow z}^{(y,z)})}{\text{var}(\Phi_{\downarrow z}^{(y,z)})}. \quad (3.11)
 \end{aligned}$$

The covariances are independent of sensor noise, as discussed in the previous section. However, the denominators with the variances of all

sensors along the calibration path depend on sensor noise, as denoted in (3.8). Every additional sensor in the calibration path, introduces an additional term in the denominator, which increases the bias towards zero. The total introduced bias is given by

$$\frac{\text{var}(\Phi_{\downarrow a'}^{(u,a')})}{\text{var}(\Phi_{\downarrow a}^{(u,a)})} \cdots \frac{\text{var}(\Phi_{\downarrow y'}^{(x,y')})}{\text{var}(\Phi_{\downarrow y}^{(x,y)})} \cdot \frac{\text{var}(\Phi_{\downarrow z'}^{(y,z')})}{\text{var}(\Phi_{\downarrow z}^{(y,z)})}, \quad (3.12)$$

where $\Phi_{\downarrow a'}^{(u,a')}$ denotes the noise-free measurements of sensor a . In the following, we show that *GMR*, in contrast to *OLS*, does not accumulate errors over multiple hops under certain conditions.

***GMR* multi-hop error accumulation.** We use the line topology above to show that, in contrast to *OLS* fitting, *GMR* does not suffer from error accumulation if all rendezvous sets of a sensor have the same variance. For example, we require that the measurements of sensor y in Figure 3.5 have the same variance in the rendezvous involving sensor x and sensor z , i.e., $\text{var}(\Phi_{\downarrow y}^{(x,y)}) = \text{var}(\Phi_{\downarrow y}^{(y,z)})$. We prove our statement by showing that the *GMR* calibration of a sensor is independent of the calibration path. Later, we release the above assumption of equal variances and analyze error propagation in the general case. We show that the error obtained is distinctly lower than with *OLS*.

Theorem 1. *Let us consider a rendezvous connection graph Γ with a calibration path $u \rightarrow a \rightarrow \dots \rightarrow x \rightarrow y \rightarrow z$. Let u be a reference sensor. If $\text{var}(\Phi_{\downarrow r}^{(p,r)}) = \text{var}(\Phi_{\downarrow r}^{(r,q)})$, $\forall p, q \in N(r)$, $\forall r \in U$, then the calibration of sensor z is independent of the path between reference sensor u and sensor z .*

Proof. It suffices to show that the statement of the theorem holds for the slope β of the calibration curve, since it uniquely determines the intercept α of the calibration curve. We use (3.2) and (3.10) to derive the calibration slope $\beta_{GMR}^{u \rightarrow z}$ of sensor z , calibrated over multiple hops starting from reference sensor u :

$$\begin{aligned} \beta_{GMR}^{u \rightarrow z} &= \left(\frac{\text{var}(\widehat{\Phi}_{\downarrow y}^{(y,z)})}{\text{var}(\Phi_{\downarrow z}^{(y,z)})} \right)^{\frac{1}{2}} \\ &= \left(\frac{\text{var}(\beta_{GMR}^{u \rightarrow y} \cdot \Phi_{\downarrow y}^{(y,z)} + \alpha_{GMR}^{u \rightarrow y})}{\text{var}(\Phi_{\downarrow z}^{(y,z)})} \right)^{\frac{1}{2}} \\ &= \beta_{GMR}^{u \rightarrow y} \cdot \left(\frac{\text{var}(\Phi_{\downarrow y}^{(y,z)})}{\text{var}(\Phi_{\downarrow z}^{(y,z)})} \right)^{\frac{1}{2}} \\ &= \dots \end{aligned}$$

$$= \left(\frac{\text{var}(\Phi_{\downarrow u}^{(u,a)})}{\text{var}(\Phi_{\downarrow a}^{(u,a)})} \cdots \frac{\text{var}(\Phi_{\downarrow x}^{(x,y)})}{\text{var}(\Phi_{\downarrow y}^{(x,y)})} \cdot \frac{\text{var}(\Phi_{\downarrow y}^{(y,z)})}{\text{var}(\Phi_{\downarrow z}^{(y,z)})} \right)^{\frac{1}{2}} \quad (3.13)$$

$$= \left(\frac{\text{var}(\Phi_{\downarrow u}^{(u,a)})}{\text{var}(\Phi_{\downarrow z}^{(y,z)})} \right)^{\frac{1}{2}}. \quad (3.14)$$

We leverage in (3.14) the assumption of equal variances, i.e., $\text{var}(\Phi_{\downarrow y}^{(x,y)}) = \text{var}(\Phi_{\downarrow y}^{(y,z)})$. Since the calibration slope $\beta_{GMR}^{u \rightarrow z}$ of sensor z only depends on the measurements of sensor u and z , i.e., on $\text{var}(\Phi_{\downarrow u}^{(u,a)})$ and $\text{var}(\Phi_{\downarrow z}^{(y,z)})$, it is independent of the choice of intermediate sensors along the calibration path. \square

If the assumption of the theorem holds, calibration parameters calculated with *GMR* are independent of the choice of calibration paths and are *optimal*, since the minimization function of *GMR* defined in (3.9) achieves its minimum.

Relaxing the assumption. The theorem assumption implies that the variance of the phenomenon signal at rendezvous between any pair of sensors must be the same. In reality this is hard to achieve, since meeting points between sensors cannot be enforced. Therefore, in many real systems the reduction used from (3.13) to (3.14) does not apply and, thus,

$$\beta_{GMR}^{u \rightarrow z} = \left(\frac{\text{var}(\Phi_{\downarrow u}^{(u,a)})}{\text{var}(\Phi_{\downarrow z}^{(y,z)})} \cdot \prod_{\substack{\forall p \rightarrow r \rightarrow q \\ p, q \in N(r)}} \frac{\text{var}(\Phi_{\downarrow r}^{(r,q)})}{\text{var}(\Phi_{\downarrow r}^{(p,r)})} \right)^{\frac{1}{2}}. \quad (3.15)$$

In multi-hop calibration, *GMR* does not continuously introduce a bias in one direction (like *OLS* towards zero), since both nominator and denominator contain the sensor noise term (see (3.8)):

$$\frac{\text{var}(\Phi_{\downarrow r}^{(r,q)})}{\text{var}(\Phi_{\downarrow r}^{(p,r)})} = \frac{\text{var}(\Phi_{\downarrow r'}^{(r',q)}) + \text{var}(e_r)}{\text{var}(\Phi_{\downarrow r'}^{(p,r')}) + \text{var}(e_r)}. \quad (3.16)$$

This observation is essential to understand the small error propagation property of *GMR*. *OLS* continuously underestimates the sensor slopes, and, hence, introduces an ever-increasing bias towards zero. In contrast, *GMR* closely approaches the true slope (sometimes underestimating and other times overestimating it) and, thus, does not accumulate a bias towards one direction.

Next, we show through extensive simulations that *GMR* achieves significantly lower calibration errors than *OLS*.

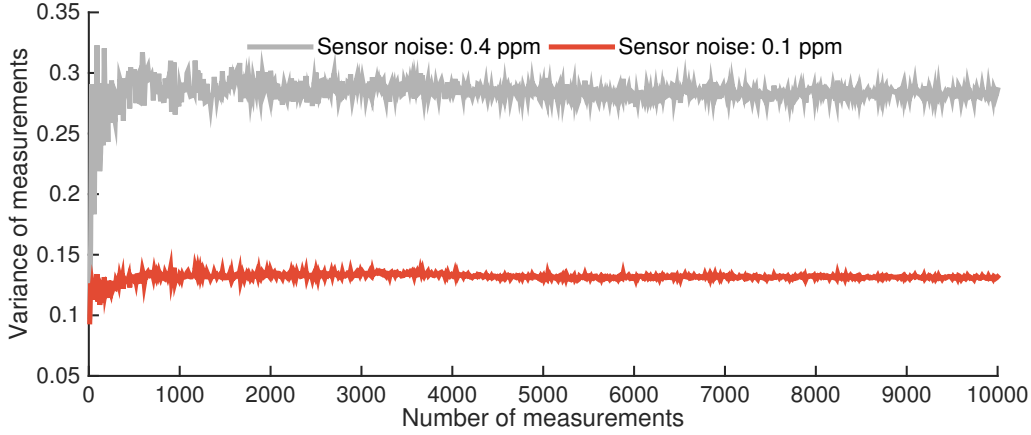


Figure 3.6: Impact of measurement quantity and sensor noise on the variance estimate. High noise and low number of measurements impede the accurate estimation of the variance.

3.5.2 Simulation Results

In the following, we use a line topology graph Γ and artificially generated rendezvous Φ between pairs of adjacent sensors to highlight the differences between *OLS* and *GMR*. Generated data correspond to typical carbon monoxide (CO) concentrations in urban environments. Artificial data generation gives us three advantages: we (i) can freely choose the topology of the rendezvous connection graph, (ii) can precisely control sensor characteristics (e.g., noise level, calibration parameters), and (iii) know for every generated sensor reading the true phenomenon signal value, which allows us to accurately evaluate calibration errors.

Setup. All generated line topologies comprise 21 sensors, with a noise-free reference sensor at the beginning of the line, followed by 20 noisy sensors. The sensor IDs denote the hop distance to the reference sensor. A rendezvous $\Phi^{(u,u+1)}$ between a pair of sensors u and $(u + 1)$ in Γ is constructed by uniformly sampling 1000 phenomenon signal values in the range $[0.2, 2]$ ppm. This range is typical for carbon monoxide concentrations in urban areas.

We randomly sample for each sensor u the calibration parameters $\alpha_u \in [0, 1]$ and $\beta_u \in [0.3, 2]$. We use these parameters to convert the generated phenomenon signal values to sensor-specific raw measurements and add sensor noise e_u to get uncalibrated sensor measurements m_u . The variance of the sensor noise is randomly sampled from $[0, 0.2]$ ppm for each sensor. The RMSE between phenomenon signal and calibrated sensor measurements is used to quantify calibration errors. All plots show average RMSEs over 100 randomly generated line topologies. For

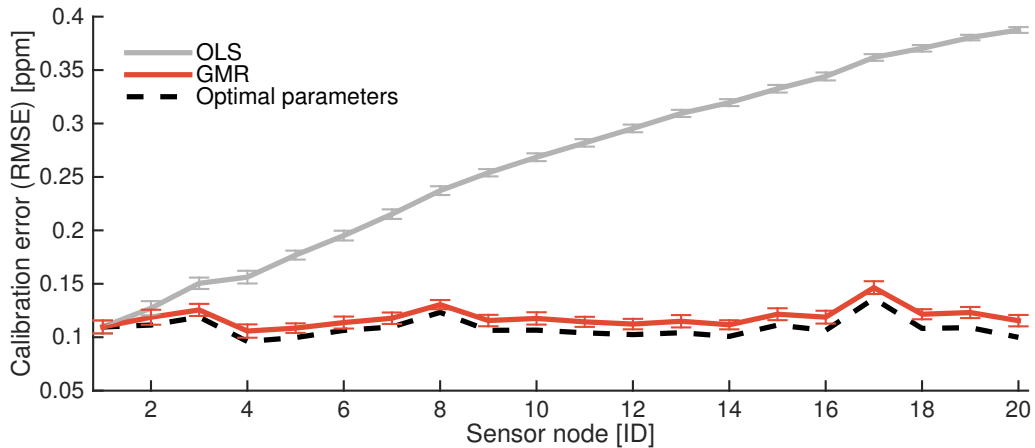


Figure 3.7: Fixed phenomenon range (Theorem 1 holds). *GMR* can accurately calibrate all sensor nodes of a 21-node line topology with very small error propagation.

each sensor, *optimal parameters* are obtained by computing *OLS* regression between raw sensor data (including noise) and the known phenomenon signal. The optimal parameters represent the best calibration parameters achieving the minimum possible RMSE between phenomenon signal and calibrated sensor measurements.

Impact of sensor noise. We need to know the variances of the sensor measurements to compute the sensors' calibration slopes, as described in the previous section. We estimate the variances based on rendezvous Φ between pairs of sensors. However, both (i) low number of measurements in rendezvous and (ii) sensor noise make an accurate variance estimation difficult. We exemplify this in Figure 3.6 using artificial data. We show the fluctuation of the variance estimates for two noisy measurement sets depending on the number of measurements in the set. As expected, variance estimation exhibits less fluctuation as the number of measurements increases. Moreover, we observe that higher sensor noise leads to higher variance estimation errors.

In Figure 3.7, we compare the calibration errors of *OLS* and *GMR* assuming that the values of the phenomenon signal for all rendezvous are sampled from the same fixed interval (the statement of Theorem 1 holds). Since the assumption of Theorem 1 holds, in theory *GMR* should not accumulate any calibration errors. With noisy sensor data, however, we observe an insignificant error accumulation due to small inaccuracies in estimating the variance of the measurements (for node 20 there is a small difference between optimal parameters and *GMR*). Nevertheless, the RMSE of *GMR* is close to the RMSE achieved with optimal parameters.

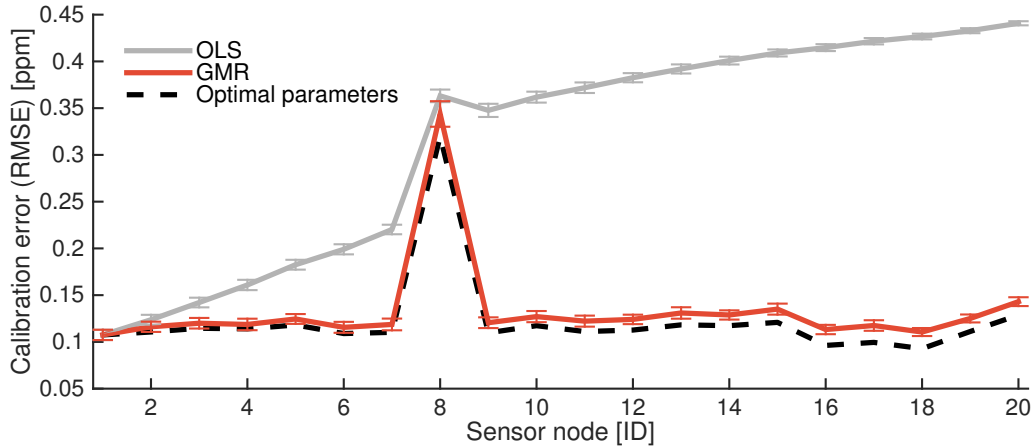


Figure 3.8: Despite the 2x higher noise of sensor 8, *GMR* accurately calibrates sensors 9–20. With *GMR* a sensor’s calibration is mostly independent of the calibration parent’s sensor noise.

In contrast, *OLS* suffers from rapid error accumulation. This is the result of regression dilution: due to sensor noise, the calibration slope of a sensor has a stronger bias towards zero with every additional calibration hop.

We make the impact of noise on the calibration error more explicit by increasing the noise of sensor 8 by factor two. Naturally, sensor 8 gets worse RMSE for any parameter setting, as depicted in Figure 3.8. We observe that *OLS* suffers from high error increase for nodes 9–20, due to the high bias towards zero of the calibration slope of sensor 8. In contrast, the nodes calibrated by *GMR* are mostly independent of the high noise of sensor 8. There is a minor error increase due to an increased inaccuracy in estimating the variance $\text{var}(\Phi_{\downarrow 8}^{(8,9)})$ of the measurements of sensor 8.

The above analysis shows that for *GMR* (i) calibration error of a sensor mostly depends on the sensor’s noise, but the noise of the calibration parent has minor impact, (ii) calibration error of a sensor can be smaller than the calibration error of its calibration parent, and (iii) there is almost no multi-hop accumulation of errors. The last two findings are crucial when calibrating a network of heterogeneous sensors with different noise characteristics.

Impact of phenomenon signal range. The assumption of Theorem 1 rarely holds for real data sets, because rendezvous between sensor nodes might occur at different ranges of the phenomenon signal. In the following, we use the line topology above but randomly choose for each rendezvous the maximum signal range in $[1, 4]$ ppm. The resulting RMSE of each sensor is plotted in Figure 3.9. Both fitting methods suffer from error accumulation as showed by the theoretical analysis, although

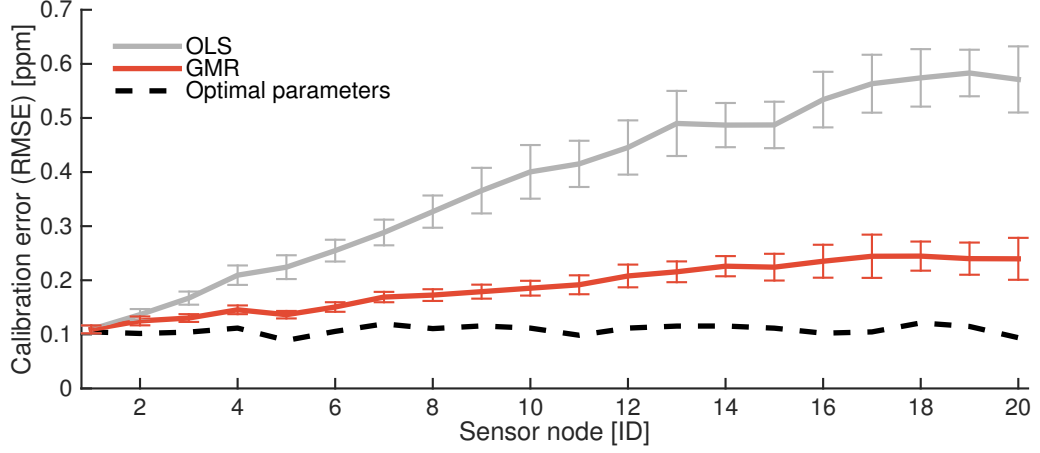


Figure 3.9: Variable phenomenon range (Theorem 1 does not hold). *GMR* has distinctly slower error accumulation than *OLS* when calibrating sensors on a 21-node line topology.

GMR is less affected than *OLS*.

To better understand the source of error accumulation, we use equal phenomenon signal ranges for all rendezvous, except between sensor 8 and 9, i.e., $\widehat{\Phi}^{(8,9)}$. The latter range is decreased in Figure 3.10 and increased in Figure 3.11 by factor three compared to the rendezvous of all other sensors. This changes the respective *range-to-noise ratios* (*RNRs*). Decreasing the range lowers *RNRs*, which means that the noise has an increasing impact on the sensor measurements. This, in turn, makes it more difficult to estimate the variance of the measurements (see Figure 3.6) and, hence, results in a higher calibration error of *OLS* and *GMR*. Moreover, the introduced error affects any sensor that has sensor 9 in its calibration path, as (3.11) and (3.15) suggest and Figure 3.10 shows. By contrast, increasing the range, raises the *RNR*. The calibration error of both fitting methods increases because the calibration of sensor 8 is based on measurements from the standard (small) range. However, in the rendezvous with sensor 9, sensor 8 has to measure in a much larger (factor three) range. This introduces with *GMR* only a small error, because *GMR* can accurately estimate the true slope of sensor 8, even if it only has a small range of measurements available. However, the bias introduced by *OLS* has a large impact on the calibration error when the range is increased. Hence, increasing the range has a much larger impact on its calibration accuracy, as shown in Figure 3.11.

We conclude that (i) varying the phenomenon signal range between rendezvous degrades the performance of both *GMR* and *OLS*, although *OLS* is significantly more affected than *GMR*, and (ii) a low *RNR*

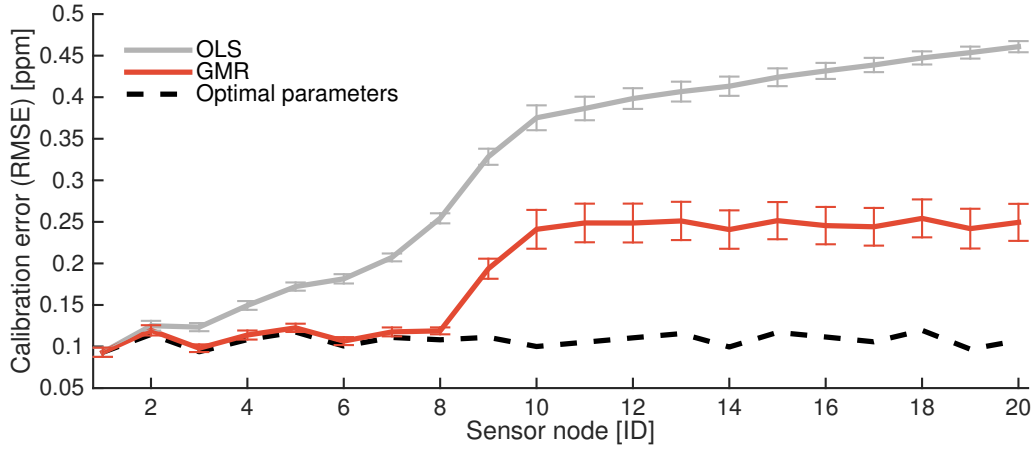


Figure 3.10: The 3x smaller measurement range between sensor 8 and 9 results in lower range-to-noise ratios of the two sensors, which influences the calibration error of all subsequent sensors.

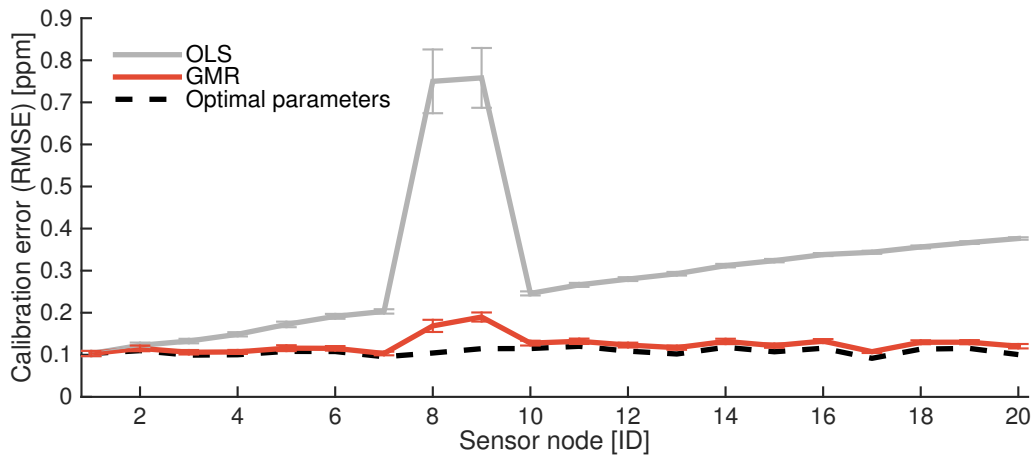


Figure 3.11: The 3x larger measurement range between sensor 8 and 9 results in higher range-to-noise ratios of the two sensors. This does not influence the calibration of sensors 10–20.

is particularly problematic, since the introduced calibration error propagates over multiple hops. Our findings suggest that calibration parents need to be carefully selected to minimize error propagation in the network.

3.6 Network Calibration

In this section, we present our calibration parent selection algorithm and analyze it in combination with the *OLS* and *GMR* line fitting methods on artificial topologies. Later, we evaluate their performance on real air pollution data sets.

3.6.1 Selecting the Calibration Parent

We showed with our previous analysis that an efficient parent selection algorithm must build on the following two properties:

- Maximize number of checkpoints with calibration parents to reduce inaccuracies in estimating the variance.
- Avoid calibration parents with low *RNR* to reduce multi-hop error propagation.

Optimizing each of the above properties is not straightforward for the following reasons: Given a noisy measurement, it is impossible to split it into its noise-free and noise components, as done in (3.8), without additional knowledge about the sensor noise. Therefore, it is difficult to reliably identify parent candidates with low *RNR*. Similarly, despite high number of calibration pairs, we may have low correlation between paired values leading to a poor regression line. In the worst case, the phenomenon signal has the same value for all calibration pairs making them useless for sensor calibration.

In order to achieve the properties listed above, we propose to (i) use a combination of multiple calibrated sensors as calibration parent to increase the number of measurements involved in sensor calibration, and (ii) analyze the Pearson correlation with the calibration parent to avoid sensors with low *RNR*.

Combining rendezvous of multiple calibrated sensors u, \dots, v and using the combination as calibration parent results in the following slope computation:

$$\beta_{GMR}^{u, \dots, v \rightarrow z} = \left(\frac{\text{var}(\widehat{\Phi}_{\downarrow u}^{(u,z)} \cup \dots \cup \widehat{\Phi}_{\downarrow v}^{(v,z)})}{\text{var}(\widehat{\Phi}_{\downarrow z}^{(u,z)} \cup \dots \cup \widehat{\Phi}_{\downarrow z}^{(v,z)})} \right)^{\frac{1}{2}}. \quad (3.17)$$

Choosing multiple parents (*i*) maximizes the number of measurements used to compute the calibration slope, and (*ii*) potentially increases the phenomenon signal range and, thus, increases the *RNR*. The latter holds with high probability if rendezvous occur randomly at distinct values of the phenomenon signal. *RNR* is then maximized as the result of the increased phenomenon range.

To minimize the chance of having a parent with low *RNR*, we compute the Pearson correlation between measurements of different sensors, i.e., between measurements in $\Phi_{\downarrow z}^{(u,z)} \cup \dots \cup \Phi_{\downarrow z}^{(v,z)}$ and $\widehat{\Phi}_{\downarrow u}^{(u,z)} \cup \dots \cup \widehat{\Phi}_{\downarrow v}^{(v,z)}$ of (3.17), and use it as noise estimate. We *blacklist* rendezvous between sensors if their correlation is below a given threshold.

Putting everything together. The last key piece is to tie all parts together. Calibration parameters may vary over time, since low-cost sensors typically loose sensitivity over time and are influenced by environmental conditions. This has significant influence on the sensor measurements and results in the fluctuation of the calibration coefficients α and β (see Figure 3.1). We introduce a periodic calibration approach by partitioning the time domain T into a sequence of equally-spaced, non-overlapping time intervals $\{s_k\} = S \subseteq T$. We assume that the calibration parameters of all sensors do not change within s_k . For each s_k , we consider all meeting points that occurred between pairs of sensors and construct a rendezvous connection graph Γ_{s_k} , which is used as input for the calibration algorithm.

Using the rendezvous connection graph Γ_{s_k} , the calibration algorithm calibrates the network of noisy sensors hop-by-hop starting from a reference sensor. The calibration algorithm starts by calibrating sensors that directly pass by reference sensors (first-hop sensors). Once calibrated, the sensors can be used as references by the second-hop sensors. The algorithm stops as soon as all sensors in the network are calibrated. We use for the calibration of a sensor $u \in U \setminus U^*$, all non-blacklisted previously calibrated sensors in $N(u)$ as calibration parent.

3.6.2 Simulation Results

We end this section by showing a comparison between hop-by-hop calibration with *OLS* and *GMR* and demonstrating the benefit of our parent selection strategy. We generate artificial data as described in the previous section, but use a grid topology in order to have multiple possible calibration parents for every sensor node. We arrange 36 nodes on a grid with 35 uncalibrated sensors and one reference node in the top left corner of the grid. The node IDs increase with increasing hop-distance from the reference sensor, i.e., sensor 1 and 2 are the direct neighbors of the reference sensor.

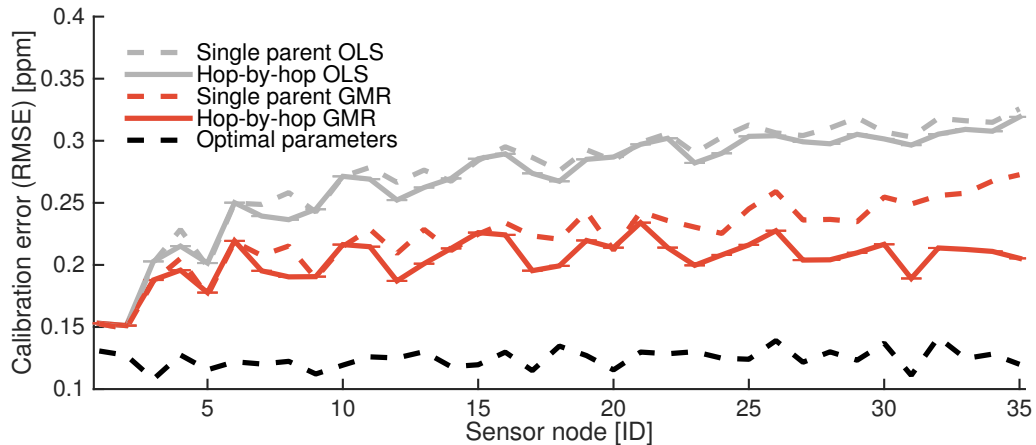


Figure 3.12: Comparison of *OLS* and *GMR* on a 35-node grid topology. The hop-by-hop calibration algorithms use multiple calibration parents and, thus, achieve a lower calibration error than the single parent version.

Figure 3.12 depicts the comparison between two versions of *OLS* and *GMR*: *hop-by-hop*, using the parent selection scheme described above and *single parent*, selecting as calibration parent the sensor with the maximum number of measurement pairs. We see again that *OLS* accumulates errors with a much faster rate than *GMR* and that multiple parents help to reduce the calibration error.

3.7 Calibrating the Mobile Sensor Network

In this section, we use the proposed network calibration algorithm to calibrate a large set of air pollution measurements collected with our mobile sensor nodes installed on top of ten streetcars of the public transport network, as described in Section 2.3.1. We demonstrate on real measurements that Hop-by-hop *GMR* considerably reduces error propagation and is able to more accurately calibrate a set of low-cost gas sensors than Hop-by-hop *OLS*.

In this chapter, we focus on the calibration of the temperature, ozone (O_3), and carbon monoxide (CO) data sets. We evaluate 13 million measurements collected in six months between March and August 2014, as summarized in Table 3.1.

Two stations of the local governmental measurement network are located in the city center very close to the streetcar tracks, as depicted in Figure 2.8 in Chapter 2. We use the stations' high-quality measurements to calibrate our low-cost sensors deployed on top of the streetcars.

Data set	Measurements [in millions]	Sampling interval [s]
Temperature	2.7	20
Ozone (O ₃)	2.1	20
Carbon monoxide (CO)	8.5	10

Table 3.1: Evaluated data sets of the mobile air pollution monitoring network collected from March to August 2014.

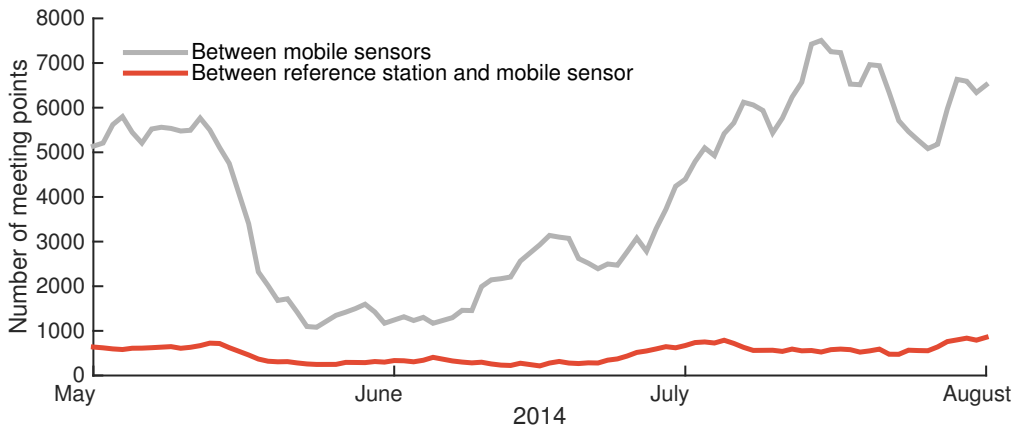


Figure 3.13: Number of meeting points among mobile sensors and reference stations accumulated every ten days during three months of operation.

3.7.1 Setup

We use all measurements to create every ten days a rendezvous connection graph. We assume that the measurements from two different sensors are similar if their spatial and temporal distances are within $\Delta d = 50$ m and $\Delta t = 5$ min. We found through extensive evaluations that this parameterization achieves a good trade-off between number of meeting points and similarity of the measured phenomena. The automatic selection of the most meaningful parameter settings has yet to be investigated. On average, within a slot size of ten days, the rendezvous connection graph of the ozone measurements comprises 4000 meeting points among mobile sensors and 500 meeting points between reference stations and mobile sensors, as shown in Figure 3.13. We blacklist a rendezvous if the correlation of the sensors' measurements is below 0.5.

3.7.2 Evaluation

We evaluate the calibration accuracy with a two-fold cross-validation approach. We calibrate the network using one reference station and evaluate the calibration accuracy with the second reference station. In

Data set	Average calibration error [RMSE]	
	Single calibration	Periodic calibration
Temperature	2.3 °C	1.6 °C
Ozone (O ₃)	12.9 ppb	9.8 ppb
Carbon monoxide (CO)	0.23 ppm	0.08 ppm

Table 3.2: Average calibration errors. Single calibration, only calibrates sensors once at the beginning of the deployment, compared to periodic calibration, which calibrates the sensors every ten days.

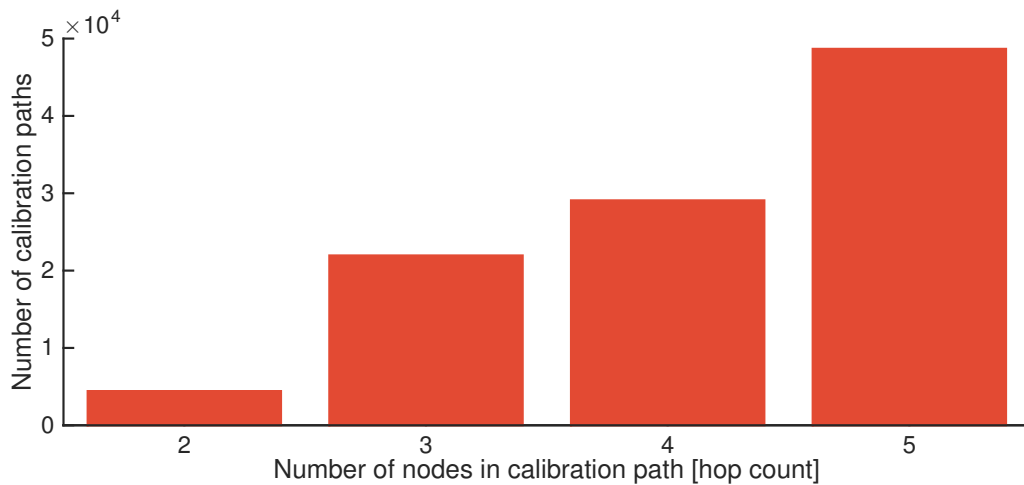


Figure 3.14: Number of calibration paths with two to five sensor nodes.

every time slot, we repeat this approach with both reference stations acting once as reference station. Table 3.2 shows the average RMSE between calibrated measurements of the mobile sensor nodes and the reference station. The achieved accuracy is reasonable for the deployed low-cost sensors and is considerably better than what is obtained if the sensors are only calibrated once at the beginning of the deployment, also listed in Table 3.2. Different applications require different accuracy levels, e.g., detecting whether the concentration is below or above a given threshold is easier than delivering accurate absolute measurements. As a rule of thumb, as long as the calibration error is below 20% in comparison to the intraurban concentration variation (as found in our case), the measurements can be used to analyze the spatial distribution of air pollutants.

We illustrate the advantages of *GMR* by extracting specific calibration path lengths from the rendezvous connection graph. We compare the calibration errors of *GMR* and *OLS* by analyzing more than 100,000 different calibration paths composed of 2 to 5 sensor nodes, as shown

Data set	Sensors calibrated		Total sensors
	one-hop	multi-hop	
Temperature	4.7	8.2	9
Ozone (O ₃)	3.8	6.2	7
Carbon monoxide (CO)	4.5	8.2	10

Table 3.3: Average number of sensors calibrated over a 10-day period using one-hop and multi-hop calibration algorithms.

in Figure 3.14. We show in Figure 3.15 that on all three pollution data sets the error accumulation of *GMR* is considerably lower than with *OLS*. The difference between *OLS* and *GMR* increases for longer path lengths. These results with real measurements are in line with our previous findings based on theoretical analysis and simulations.

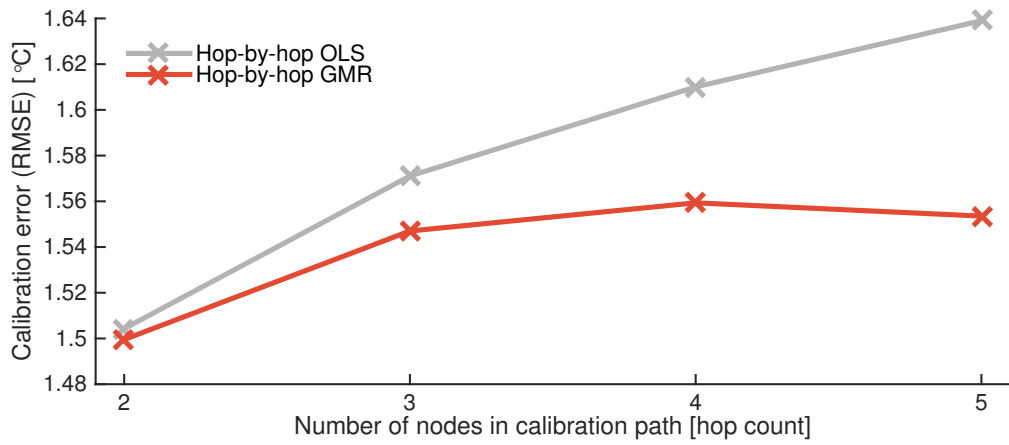
Finally, we compare in Table 3.3 the number of sensors, which can be calibrated with existing one-hop algorithms (i.e., only calibrating sensors directly passing by a reference sensor, e.g., [MLCOS08]) to the number of nodes reached with our multi-hop calibration approach. Note that not all ten sensor nodes are equipped with all sensor types. We observe that our multi-hop algorithm calibrates at least 34 % more sensors than its one-hop alternative. This is remarkable, since our air pollution monitoring network is well-connected with two reference stations located in the city center. We expect an even higher benefit when using multi-hop calibration algorithms for large-scale systems, e.g., in participatory sensing applications.

3.8 Related Work

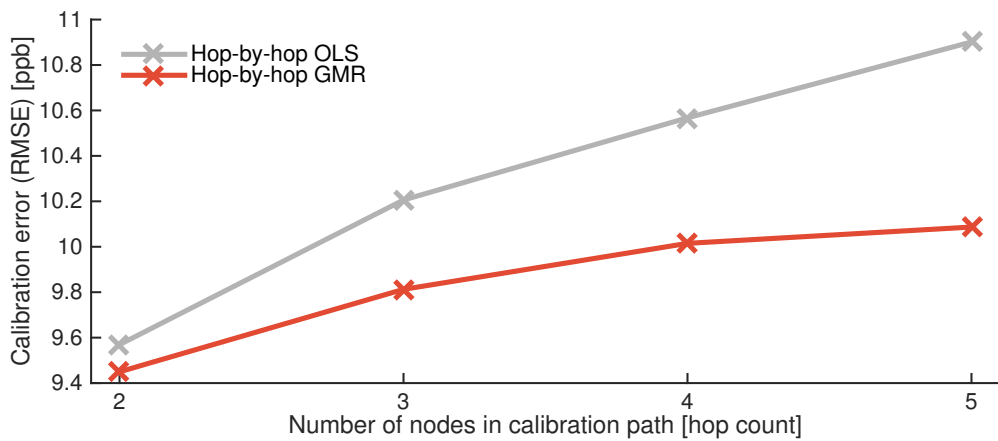
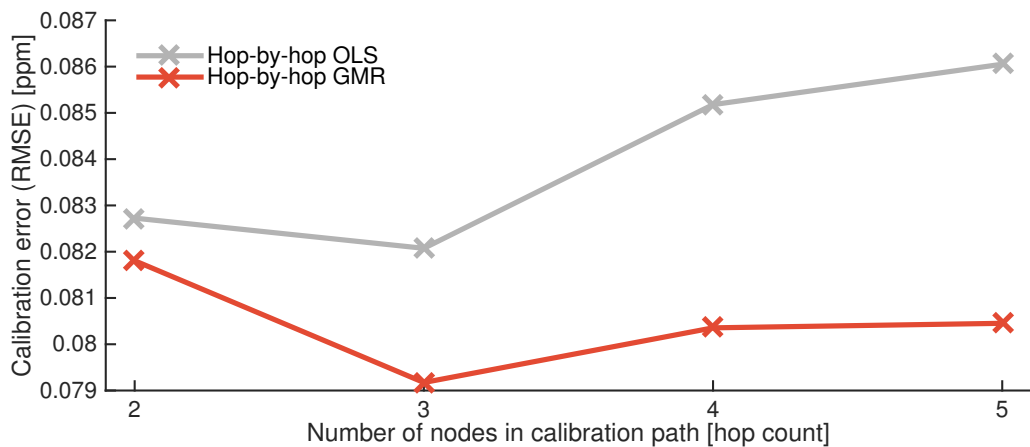
This section summarizes existing knowledge related to this chapter. We overview line fitting methods in the context of sensor calibration, list state-of-the-art sensor calibration algorithms, and discuss their differences to the work described in this chapter.

3.8.1 Line Fitting

Linear regression is a common statistical operation in many sciences. Besides of *OLS* and *GMR*, both discussed in detail in this work, there is a broad range of line fitting methods, such as Theil-Sen estimator [The50, Sen68], which chooses the median slope among all lines through pairs of points, total least squares regression [MVH07], which optimizes the sum of squared orthogonal residuals to the regression line, and Deming regression [Dem44], which is the maximum likelihood estimator. Each



(a) Temperature.

(b) Ozone (O₃).

(c) Carbon monoxide (CO).

Figure 3.15: Dependency between calibration error and length of the calibration path. Hop-by-hop GMR has a considerably lower error accumulation than OLS for all three pollutants.

method was designed for a specific regression model, that must match application requirements. In the context of sensor network calibration, the following requirements are important: we need to (i) account for errors in both variables, (ii) handle both variables symmetrically, and (iii) be invariant to changes in the scale. Our choice of line fitting method and its alternatives are detailed below.

When calibrating a sensor using another noisy sensor, the chosen regression method must account for measurement errors in both variables. Standard *OLS* ignores the problem and thus suffers a noise-driven bias in estimating the slope, known as regression dilution [CRSC06]. The bias can be compensated [Gil06, FT00], but requires additional knowledge about the errors, e.g., their variances or the ratio thereof [Dem44]. This knowledge is often not available, especially if sensor noise depends on environmental conditions.

Many regression methods make a distinction between the predictor (independent) variable and the response (dependent) variable, because their optimization functions are asymmetric with respect to the two variables, e.g., *OLS* and Theil-Sen estimator. Thus, these regression methods derive two different regression lines depending on the direction of regression. Using asymmetric regression methods poses a dilemma in many applications, when trying to identify a cause-effect relationship between the variables, e.g., if one would want to find the dependency between body length and body mass [Smi09]. Thus, it is often desirable to have one regression line describing the relationship rather than two. A similar problem arises in the context of sensor calibration: It is hard to fix the direction of calibration for a pair of noisy sensors of the same type. In this case, symmetric regression models, such as *GMR* and total least squares, produce a more comprehensible fit.

Lack of scale invariance makes the regression dependent on the range of raw measurements limiting its applicability for sensor calibration. For example, total least squares is not scale-invariant.

GMR is known under different names such as reduced major axis [Smi09], line of organic correlation [KH50], and Strömberg's impartial line [FB92]. It is the only linear regression method in the two-dimensional space, which is simple (i.e., only based on the ratio of the variables' variances), symmetric, and scale-invariant [Sam42].

3.8.2 Sensor Calibration

A great body of work is dedicated to calibrating low-cost sensors, e.g., [DVMP⁺08, MLB⁺12, MLCOS08, LB14]. The existing literature on sensor calibration can be classified into micro- and macro-calibration approaches [WC02, TXY⁺13].

Micro-calibration. Micro-calibration algorithms calibrate every sensor in the network according to a high-quality reference signal [TPS⁺05, RBB⁺06, MLCOS08, HST12]. The goal is to have accurate absolute sensor measurements of the phenomenon monitored.

Tolle et al. [TPS⁺05] and Ramanathan et al. [RBB⁺06] calibrate their sensors in the laboratory before the deployment phase. Manual calibration is an elaborate and time-consuming task and many sensors need periodic re-calibrations to deliver measurements with a good data quality during the whole system lifetime (see Table 3.2).

Miluzzo et al. [MLCOS08] propose CaliBree, a self-calibration system for mobile sensor nodes. CaliBree makes use of meeting points between low-cost sensors and high-quality reference sensors to update the calibration of low-cost sensors. Compared to the algorithm proposed in this chapter, it ignores all meeting points between low-cost sensors and, thus, requires in general a high density of reference sensors to calibrate all sensors in a network (see Table 3.3).

Hasenfratz et al. [HST12] introduce a multi-hop calibration algorithm applying *OLS* regression to calibrate a network of low-cost sensors. They show with simulations a linear increase of calibration error with increasing number of nodes in the calibration path. We show with our theoretical analysis, simulations, and real pollution measurements a considerably smaller error accumulation.

Macro-calibration. Macro-calibration algorithms try to maximize the similarity among the measurements of all sensor nodes in the network [BMEP03, FMP03, BN07, LB14, TXY⁺13]. In macro-calibration, the main goal is not to adjust sensor calibration according to a reference signal but to achieve a high similarity between sensors in the network. Hence, most macro-calibration algorithms do not require access to reference measurements.

Bychkovskiy et al. [BMEP03] propose a two phase multi-hop calibration algorithm. In the first phase, the algorithm derives relationships between co-located sensors and in the second phase it maximizes the consistency among groups of sensor nodes. The authors assume that sensor noise can be filtered out. In general, this is very difficult to achieve in mobile sensing networks.

Balzano and Nowak [BN07] and Lipor and Balzano [LB14] introduce blind calibration algorithms to determine sensor gains by oversampling the phenomenon signal. The authors compare different methods to derive calibration parameters: singular value decomposition and *OLS* in [BN07], and total least squares in [LB14]. Blind calibration methods depend on the assumption that the mean of the monitored signal is zero (or the same for all sensors). This is in most real-world deployments not the case, e.g.,

air pollution monitoring.

Xiang et al. [XBP⁺12] propose a collaborative sensor calibration and placement approach using meeting points between mobile sensors to adjust their calibration parameters and error estimates. The goal is to compensate for sensor drift errors. The measurements from our long-term, mobile air quality monitoring system indicate that sensor drift has a minor effect on the fluctuation of calibration parameters. In contrast, dynamically changing environmental conditions highly affect sensor performance. This observation is in line with the findings published in [RBB⁺06].

Other macro-calibration algorithms improve target detection [TXY⁺13], localization [WC02], and the calibration of light sensors [FMP03].

In contrast to all above works, the distributed monitoring of physical phenomena (e.g., urban noise monitoring [MPSS13], earthquake detection [FOC⁺11], and air pollution monitoring [RPV14]) requires accurate absolute sensor measurements. Hence, in these scenarios a micro-calibration approach is needed, such as the calibration algorithm proposed in this chapter. Moreover, a good micro-calibration algorithm also accomplishes the goals of macro-calibration, namely a high consistency among measurements of co-located sensor nodes, whereas the opposite is not true.

3.9 Conclusions

Wireless sensor networks (WSNs) are used in an increasing number of applications enabling the dense monitoring of the environment. Many sensors used in WSNs need to be frequently calibrated to constantly deliver accurate measurements.

Calibrating a whole network of sensors based on a few high-quality reference measurements is challenging. Error propagation in the network hinders the accurate calibration of all sensor nodes. In this chapter, we propose a novel multi-hop calibration algorithm, which distinctly reduces error accumulation in the network. Instead of using ordinary least squares—typically employed to calibrate noisy sensors—we use geometric mean regression and exploit its robustness against regression dilution, which is caused by noisy sensor readings. This is essential in order to reduce error propagation in multi-hop calibration systems.

We show the benefits of our algorithm by performing a detailed theoretical analysis, conducting simulations under various settings, and calibrating measurements from our mobile air pollution monitoring

network. We demonstrate the following key features of our calibration algorithm: (i) Very low hop-by-hop error propagation in the network. Hence, the algorithm can handle long calibration paths incorporating many sensors without accumulating large errors. (ii) The calibration quality of a sensor is largely independent of the calibration error of its calibration parent. A low-noise sensor can be accurately calibrated despite having a high-noise calibration parent. (iii) Due to multi-hop calibration, only a low number of reference sensors are required since meeting points between low-cost sensors are also exploited as calibration opportunities.

Hence, the algorithm proposed is suitable for the accurate calibration of large, mobile, and heterogeneous sensor networks, such as found in participatory sensing scenarios. Further, it enables the accurate calibration of our mobile air pollution monitoring network.

In Chapter 4, we introduce a model-driven approach to provide individual accuracy bounds for calibrated sensor measurements and to detect systematic and transient sensor errors.

4

Accuracy Bounds for Low-Cost Sensor Readings

Wireless sensor networks are increasingly used in application scenarios where a high data quality is inevitable. Nevertheless, many deployments must live with strict constraints regarding the sensing hardware and may have to employ sensors with limited reliability and accuracy, e.g., due to limited energy budget, size, and bandwidth. Additionally, many applications would benefit from not only gathering absolute sensor readings but also knowing the quality of their low-cost sensor measurements, e.g., to ignore inaccurate sensor readings and to schedule extra measurements. In this chapter, we use calibrated sensor readings and introduce a model-driven approach that (i) provides reliable accuracy bounds for individual noisy measurements and (ii) detects systematic and transient sensor errors. We apply our method to temperature and ozone data sets collected with our static and mobile deployments. We find that the proposed algorithm successfully calculates precise accuracy bounds. We compare them to measurements of high-quality instruments and show that up to 96% of the reference measurements are inside the computed accuracy bounds in the static scenario and up to 94% in the mobile scenario. By analyzing data from our static long-term deployment, we reveal that the ozone sensor's reliability is dependent on seasonal weather conditions.

4.1 Introduction

Until a few years ago wireless sensor networks (WSNs) were mainly deployed by system researchers to understand and solve fundamental challenges (e.g., time synchronization and reliable communication). Examples are short-living deployments, such as structural monitoring for a few days [XRC⁺04] and habitat monitoring for several months [TPS⁺05]. The progress in the last years enabled researchers to install long-term deployments operating for multiple years. Examples of long-standing installations are the monitoring of permafrost in alpine regions [BGH⁺09, KWL⁺11] and heritage buildings in Northern Italy [CMP⁺09], both in operation for multiple years. The improvements in wireless technologies also made it possible to install large-scale WSNs in industrial production areas [GH09, PSLN⁺12]. In many of these recent deployments the quality of the gathered data is vital for the success of the whole application. Nowadays, data collected by WSNs are used for adaptive lighting in road tunnels [CCD⁺11], data center monitoring [LLL⁺09], clinical patient surveillance [CLBR10], and, as in our case, for pollution monitoring [RPV14]. In all these application scenarios wrong or inaccurate sensing may lead to wrong decisions with crucial impact on the failure safety of the system.

Challenges. Many WSN deployments have strict constraints regarding the sensing hardware. The deployed sensors need to be low-cost in multiple domains, such as size, energy consumption, and price. The strict obedience to these design goals typically leads to a *reduced measurement accuracy and reliability* of the produced sensors [EN03a, MSPD09]. One common example are sensors for air quality monitoring. Fixed stations use state-of-the-art instruments, which are very accurate but have a large size and cost tens of thousands of dollars. Low-cost gas sensors (~100 dollars) are compact in size and suitable for mobile measurements but show a 100x higher uncertainty than the standard instruments.

Additionally, for many applications it is relevant to not only gather the absolute sensor readings, but also to estimate the quality of individual measurements. The measurement quality depends on many factors, such as type of sensor used, sensor aging effects, calibration quality, environmental conditions, and type of the monitored phenomenon. Certain applications may only allow a small tolerance in the inaccuracy of the collected data [RCM⁺12].

Contributions and road-map. We exploit the fact that every measurable phenomenon underlies certain physical processes. This enables us to calculate upper and lower bounds on the maximum feasible phenomenon

change over a given temporal and spatial distance. Similarly, also sensors show an assessable performance over time allowing to calculate bounds on the maximum sensor drift for a given time period. We use these bounds together with the collected sensor readings to augment each measurement with individual accuracy bounds, indicating the reliability of the measured values. Our approach does not impose any overhead on the sensor nodes. Furthermore, the proposed method automatically detects and filters out systematic (e.g., hardware failure) as well as transient (e.g., outliers) sensor errors.

The main contributions described in this chapter are:

- We introduce a novel model-driven algorithm to calculate reliable accuracy bounds for noisy sensor readings and detect erroneous measurements.
- We demonstrate the applicability of our method by using the proposed algorithm to analyze large sets of static and mobile temperature and ozone measurements from our deployments described in Chapter 2.
- By comparing the sensor readings to measurements of high-quality instruments, we show that the calculated accuracy bounds are highly accurate, and by analyzing 1.5 years of data from our sensor testbed we reveal a dependency between the ozone sensor's reliability and seasonal weather conditions.

We give in Section 4.2 a system overview and introduce generic models to characterize the phenomenon of interest and the sensing hardware. We describe our method in Section 4.3, which uses these models to compute for each sensor reading corresponding accuracy bounds. Section 4.4 presents our case study and details the specific models that are used in Section 4.5 to evaluate two measurement campaigns. We monitor temperature and urban air pollution by using unstable and noisy low-cost sensors. We present results from (i) our static installation on top of a fixed station and (ii) our mobile deployment on top of streetcars of the public transport network. We show that the algorithm can calculate accurate bounds for different sensor types under static and mobile settings. When compared to high-quality instruments, in the static up to 96 % and in the mobile scenario up to 94 % of the reference measurements are inside the computed accuracy bounds. We survey related work in Section 4.6, and conclude in Section 4.7.

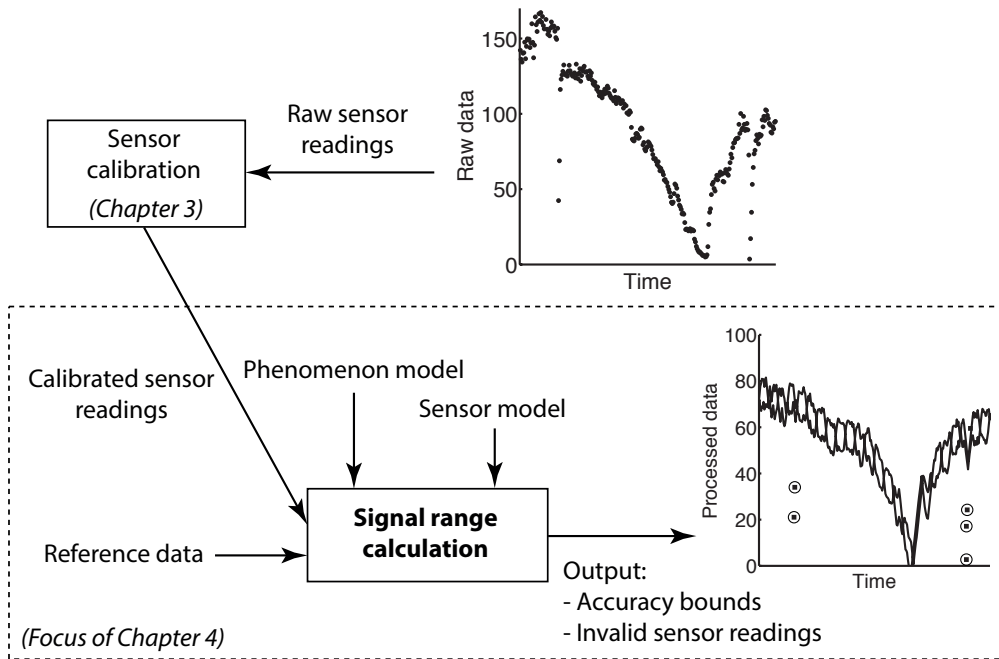


Figure 4.1: System overview. Calibrated sensor readings, reference data, and sensor and phenomenon models are used to calculate for each sensor reading individual accuracy bounds and to detect erroneous measurements.

4.2 System Models

We start by giving a high-level overview of our approach and introduce models to describe the phenomenon of interest and sensor characteristic.

4.2.1 System Overview

We start by giving a high-level overview, as illustrated in Figure 4.1. We consider application scenarios where sensor nodes monitor a specific phenomenon and transmit their measurements to a base station for further processing [BGH⁺09, CMP⁺09]. We make use of additional information at the base station consisting of (i) reference data containing at least one tuple of a reference measurement and a corresponding calibrated sensor reading collected at a rendezvous (see Section 3.3 for the definition of a rendezvous), (ii) measurements from sensor nodes in the network, and (iii) two independent models that describe the phenomenon of interest and the sensor characteristic. This information is combined to calculate for each sensor reading an upper and a lower accuracy bound indicating the quality of the measurement and to detect erroneous sensor readings. We require as input calibrated sensor readings, based on any desired calibration algorithm, e.g., [BMEP03, XBP⁺12] or the one described in

Chapter 3. We consider sensor calibration as being independent from the computation of accuracy bounds. It would be possible to combine the two, this is, however, beyond the scope of our work.

Model assumptions. We assume for the phenomenon model that the maximum phenomenon changes within a given time period and distance are bounded. Also, we assume for the sensor model that sensor drift and sensor noise are bounded. In reality, phenomenon variations and sensor characteristics are based on physical and chemical processes and are only with a certain (high) degree of probability within the assumed bounds. Hence, also the derived statements only hold with a high probability. We will show in our evaluation in Section 4.5 with real temperature and ozone measurements how the chosen model parameters impact the significance of the derived accuracy bounds.

4.2.2 Phenomenon Model

We exploit the fact that every phenomenon underlies certain physical processes. Hence, it is possible to calculate bounds on the maximum possible phenomenon change over a given temporal and spatial distance. We introduce the phenomenon model to describe the behavior of the monitored phenomenon of interest.

Model. Let a phenomenon of interest H exhibit a continuous measurable signal $h : T \times L \rightarrow D$ at time $t \in T \subseteq \mathbb{R}^+$, at location $l \in L \subseteq \mathbb{R} \times \mathbb{R}$, and with a domain of measurable values $D \subseteq \mathbb{R}$. Consider a measurement $h(t_i, l_i)$ performed by a perfect sensor at time $t_i \in T$ and location $l_i \in L$. We introduce a positive-definite function f_ω that defines the bounds on the possible signal values at time $t_j \in T$ and location $l_j \in L$ depending on the temporal and spatial distances $|t_i - t_j|$ and $|l_i - l_j|$:

$$|h(t_i, l_i) - h(t_j, l_j)| < f_\omega(|t_i - t_j|, |l_i - l_j|). \quad (4.1)$$

For simplicity, we assume that f_ω is symmetric. The generalization to an asymmetric f_ω is straightforward. The interval $[\Omega_{i,j}]^1$ represents the maximum possible signal change from time t_i and location l_i to time t_j and location l_j , that is, the *lower and upper bounds* on the maximum variation:

$$[\Omega_{i,j}] = [-f_\omega(|t_i - t_j|, |l_i - l_j|), f_\omega(|t_i - t_j|, |l_i - l_j|)]. \quad (4.2)$$

Example. Figure 4.2 shows an example where we assume, for simplicity, a static location (i.e., $l_i = l_j = l_l$). In this example, the phenomenon function

¹To make the notation of intervals more compact, we will often represent intervals with squared brackets $[a]$ without writing the exact lower and upper bounds $[a^l, a^u]$.

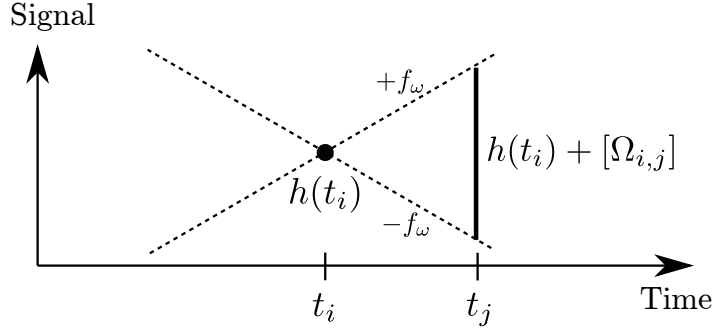


Figure 4.2: Phenomenon model. The phenomenon function f_ω is applied to signal value $h(t_i)$ at time t_i to limit possible signal values at time t_j to $h(t_i) + [\Omega_{i,j}]$.

f_ω is linear and provides bounds on the possible phenomenon change around the known phenomenon level $h(t_i)$ at time t_i . The possible signal values of the phenomenon at time t_j are bounded by $h(t_i) + [\Omega_{i,j}]$.

4.2.3 Sensor Model

We introduce the sensor model to describe the characteristic of the sensor used to monitor the phenomenon of interest. We exploit the fact that most sensors show a well-defined performance over time, e.g., solid-state gas sensors suffer from an incremental loss of sensitivity over time [HZG11].

Model. Consider a sensor that takes measurements of a phenomenon H . This leads to a sequence of discrete raw measurements $\{m(t_i)\}$ at time $t_i \in T$ (we will not explicitly write the measurement location $l_i \in L$ to keep the formulas compact). We consider a measurement as point measurement, that is, it has no duration.

If the sensor is perfect then $m(t_i) = h(t_i)$, where $h(t_i)$ is the value of the actual measured phenomenon. However, low-cost sensors typically suffer from the effects of aging, are sensitive to ambient humidity and temperature changes, and responsive to the influence of interfering gases [KBP06, HZG11]. These have a significant influence on the sensor readings and result in a deviation of $m(t_i)$ from $h(t_i)$. This deviation can be minimized by adjusting the sensor calibration, e.g., as described in Chapter 3. We denote with $\widehat{m}(t_i)$ the calibrated sensor reading.

Let us denote with $d(t_i)$ the sensor offset, that is, $d(t_i) = \widehat{m}(t_i) - h(t_i)$. The offset $d(t_i)$ is zero if the sensor is perfectly calibrated at time t_i . We introduce a monotonically increasing,² positive-definite function f_γ , which models the maximum accumulated sensor drift from time t_i to time

²The drift of most sensors is gradually increasing over time [HZG11].

$t_j \in T$ depending on the time difference $|t_i - t_j|$:

$$|d(t_i) - d(t_j)| < f_\gamma(|t_i - t_j|). \quad (4.3)$$

The interval $[\Gamma_{i,j}]$ specifies bounds on the maximum possible accumulated drift of the calibrated sensor reading from time t_i to time t_j :

$$[\Gamma_{i,j}] = [-f_\gamma(|t_i - t_j|), f_\gamma(|t_i - t_j|)]. \quad (4.4)$$

Typically, sensor readings are noisy. We describe with e_n the sensor noise and with $[\widehat{m}(t_i)] = [\widehat{m}(t_i) - e_n, \widehat{m}(t_i) + e_n]$ the calibrated noisy sensor reading.

4.3 Computation of the Signal Range

We use the introduced models to calculate accuracy bounds. For this, we use the following *interval arithmetics* given the two intervals $[a]$ and $[b]$:

$$[a] + [b] \equiv [a^l + b^l, a^u + b^u] \quad (\text{addition}) \quad (4.5)$$

$$[a] \wedge [b] \equiv [\max(a^l, b^l), \min(a^u, b^u)] \quad (\text{intersection}) \quad (4.6)$$

We define the size of an interval $[a]$ as $[\overline{a}] = a^u - a^l$, if $a^u \geq a^l$ and otherwise $[\overline{a}]$ is undefined.

4.3.1 General Approach

In the following, we introduce the concept of sensor accuracies and signal ranges, both principal elements of the algorithm introduced later on.

Sensor accuracy. No sensor is perfect, hence sensor readings have a certain uncertainty, e.g., due to sensor drift and noise. We use the sensor model introduced above to calculate for each sensor reading $\widehat{m}(t_j)$ the corresponding *sensor accuracy* $[\delta(t_j)]$. The sensor accuracy at time t_j is narrowed by the immediate preceding and succeeding sensor accuracies $[\delta(t_{j-1})]$ and $[\delta(t_{j+1})]$, and the maximum possible sensor drifts $\Gamma_{j-1,j}$ and $\Gamma_{j+1,j}$ between times $t_{j\pm 1}$ and time t_j :

$$[\delta(t_j)] = [\delta(t_j)] \wedge \bigwedge_{i=j\pm 1} ([\delta(t_i)] + [\Gamma_{i,j}]). \quad (4.7)$$

Figure 4.3 illustrates the calculation of $[\delta(t_j)]$ by using a linear sensor drift model (dotted lines). Considering only the immediately preceding and succeeding sensor readings is enough, since the sensor drift is modeled by a monotone function that is only time-dependent. The chronological order of the evaluation is irrelevant, since the interval intersection operation, as defined above, is associative.

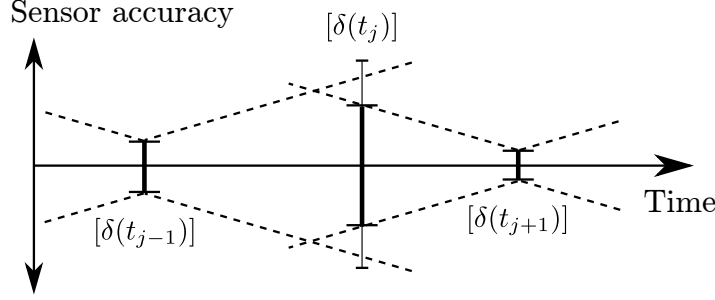


Figure 4.3: Sensor accuracy. The sensor accuracies at times $t_{j\pm 1}$ and the maximum sensor drifts (dotted lines) between $t_{j\pm 1}$ and t_j bound the sensor accuracy $[\delta(t_j)]$ at time t_j .

Without any additional knowledge, all sensor accuracies are unbounded, i.e., $[\delta(t_i)] = [-\infty, +\infty]$. Hence, we require at least one reference tuple consisting of a high-quality (i.e., higher than the low-cost sensor's quality) reference measurement $r(t_i)$ and a sensor reading $[\widehat{m}(t_i)]$, which were measured in each other's immediate spatial and temporal vicinity (e.g., as defined in Chapter 3). We assume that all reference measurements have a constant accuracy $[a_r]$. We use this additional information to adjust the sensor reading $[\widehat{m}(t_i)] = [r(t_i), r(t_i)]$ and to set the initial accuracy of this measurement to $[\delta(t_i)]^{(0)} = [a_r]$. Then, we iteratively calculate bounds for all other sensor accuracies:

$$[\delta(t_j)]^{(n+1)} = [\delta(t_j)]^{(n)} \wedge \bigwedge_{i=j\pm 1} ([\delta(t_i)]^{(n)} + [\Gamma_{i,j}]). \quad (4.8)$$

We continue with the definition of the signal range.

Signal range. We define the *signal range* $[c(t_j)]$ as the feasible phenomenon range at time t_j given the calibrated sensor reading $[\widehat{m}(t_j)]$ and sensor accuracy $[\delta(t_j)]$:

$$[c(t_j)] = [\widehat{m}(t_j)] + [\delta(t_j)]. \quad (4.9)$$

We use the phenomenon model to calculate bounds for the signal range. We reduce its size by integrating all possible phenomenon drifts $[\Omega_{i,j}]$ between signal range $[c(t_j)]$ and any other signal range $[c(t_i)]$:

$$[c(t_j)]^{(n+1)} = [c(t_j)]^{(n)} \wedge \bigwedge_{\forall i} ([c(t_i)]^{(n)} + [\Omega_{i,j}]). \quad (4.10)$$

We take all available signal ranges $[c(t_i)]$ into account (from all sensors available), since in general, the feasible change of the phenomenon is not only time- but also location-dependent. In practice, we limit the number of measurements to be evaluated by ignoring data collected outside a given radius (e.g., 1 km), as with a very high probability these measurements will not help to lower the signal ranges.

Example. We outline a simple example using temperature measurements. Consider a temperature sensor without noise (i.e., $e_n = 0^\circ\text{C}$) and with a presumed sensor accuracy of $[\delta(t_1)] = [-3, 3]^\circ\text{C}$ at time t_1 . Assume we know the sensor accuracy $[\delta(t_0)] = [-1, 1]^\circ\text{C}$ at time t_0 , which is more accurate than at time t_1 , and the sensor drift $[\Gamma_{0,1}] = [-0.5, 0.5]^\circ\text{C}$ between times t_0 and t_1 . We use this information to get tighter sensor accuracy bounds for $[\delta(t_1)]$ using (4.8):

$$[\delta(t_1)] = [-3, 3] \wedge ([-1, 1] + [-0.5, 0.5]) = [-1.5, 1.5]^\circ\text{C}. \quad (4.11)$$

Next, consider a calibrated sensor reading $\widehat{m}(t_1) = 10^\circ\text{C}$ at time t_1 . Despite that the sensor is measuring 10°C , due to the sensor's uncertainty the feasible temperature range is in $[c(t_1)] = 10 + [-1.5, 1.5] = [8.5, 11.5]^\circ\text{C}$.

We use the phenomenon model to get a tighter signal range. Consider measurement $[c(t_5)] = [12, 13]^\circ\text{C}$ and assume a maximum phenomenon change of $[\Omega_{5,1}] = [-2, 2]^\circ\text{C}$ between times t_5 and t_1 . We use (4.10) to re-evaluate $[c(t_1)]$:

$$[c(t_1)] = [8.5, 11.5] \wedge ([12, 13] + [-2, 2]) = [10, 11.5]^\circ\text{C}. \quad (4.12)$$

We curtailed the signal range from $[7, 13]^\circ\text{C}$ to $[10, 11.5]^\circ\text{C}$ by using two additional sensor readings and the bounds given by the sensor and phenomenon models.

4.3.2 Algorithm

We sketch in Alg. 1 the proposed algorithm. For the evaluation of the signal ranges, we require a finite set of calibrated sensor readings $S = \{[\widehat{m}(t_1)], \dots, [\widehat{m}(t_p)]\}$ and a finite, *non-empty* set of reference pairs $R = \{(r(t_1), [\widehat{s}(t_1)]), \dots, (r(t_m), [\widehat{s}(t_m)])\}$.

The algorithm starts with an initialization phase (lines 1–10), where all sensor readings for which we have reference measurements are adjusted and the corresponding sensor accuracies are set to $[a_r]$. The initial sensor accuracies of all other measurements are unbounded. Then, we first solve (4.8) for all uncertainties (line 14) until the sensing accuracies stay unchanged. The required number of iterations depends on the maximum number of measurements between any sensor reading $[\widehat{m}(t_j)] \in S$ and its time-wise closest reference measurement. Next, (4.10) is solved for all signal ranges (line 25) until the ranges stay unchanged. The maximum number of iteration depends on the phenomenon model used, number of invalid measurements, and their order of detection.

Detection of erroneous measurements. The intersect operation on line 25 can lead to an empty signal range, i.e., $c(t_j)^l > c(t_j)^u$. This occurs if the signal range bounds of a sensor reading are not overlapping with the

Algorithm 1 Calculation of the signal range.

Input: Calibrated sensor readings $S = \{\widehat{m}(t_1), \dots, \widehat{m}(t_p)\}$ and reference pairs $R = \{(r(t_1), \widehat{m}(t_1)), \dots, (r(t_m), \widehat{m}(t_m))\}$ with accuracy $[a_r]$.

Output: Set of signal ranges $C = \{[c(t_1)], \dots, [c(t_p)]\}$.

```

1:  $D := \emptyset$ ;
2:  $C := \emptyset$ ;
   # Initially all sensor accuracies are unbounded.
3: for all  $[\widehat{m}(t_j)] \in S$  do
4:    $[\delta(t_j)] := [-\infty, +\infty]$ ;
5:    $D := D \cup [\delta(t_j)]$ ;
6: end for
   # Reference measurements provide safe bounds.
7: for all  $(r(t_j), \widehat{m}(t_j)) \in R$  do
8:    $[\widehat{s}(t_j)] := [r(t_j), r(t_j)]$ ;
9:    $[\delta(t_j)] := [a_r]$ ;
10: end for
   # Evaluate for each sensor reading the sensor accuracy.
11: repeat
12:    $D' := D$ ;
13:   for all  $[\delta(t_j)] \in D$  do
14:      $[\delta(t_j)] := [\delta(t_j)] \wedge \bigwedge_{i=j\pm 1} ([\delta(t_i)] + [\Gamma_{i,j}])$ ;
15:   end for
16: until  $D == D'$ 
17: for all  $[\delta(t_j)] \in D$  do
18:    $[c(t_j)] := [\widehat{m}(t_j)] + [\delta(t_j)]$ ;
19:   Mark  $[c(t_j)]$  as valid;
20:    $C := C \cup [c(t_j)]$ ;
21: end for
   # Evaluate for each sensor reading the signal range.
22: repeat
23:    $C' := C$ ;
24:   for all valid  $[c(t_j)] \in C$  do
25:      $[c(t_j)] := [c(t_j)] \wedge \bigwedge_{v_i} ([c(t_i)] + [\Omega_{i,j}])$ ;
     # If the signal range is empty, build a conflict graph
     to find the erroneous measurements.
26:     if  $c(t_j)^l > c(t_j)^u$  then
27:       Build conflict graph  $G = (V, E)$ ;
28:       Find maximum independent set  $I$  of  $G$ ;
29:       Mark all  $[c(t_j)] \in V \setminus I$  as invalid;
30:     end if
31:   end for
32: until  $C == C'$ 

```

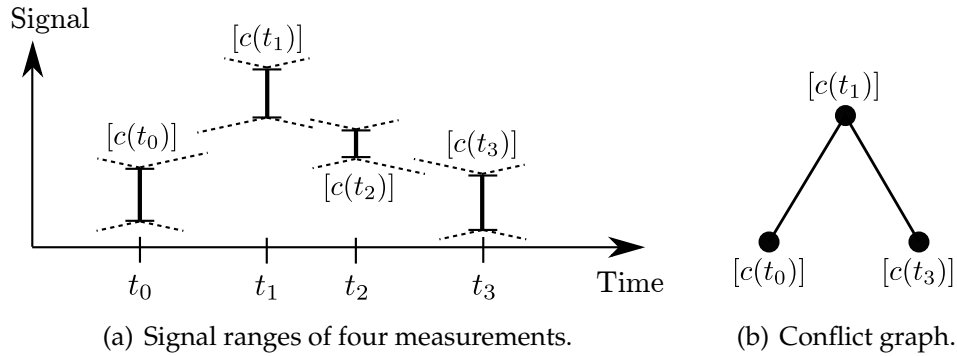


Figure 4.4: Detection of erroneous measurements. Trying to intersect (dotted lines) the four signal ranges $[c(t_0)]$ to $[c(t_3)]$ leads to several empty intervals (i.e., no intersection) in (a) resulting in the conflict graph in (b).

predicted bounds defined by the phenomenon and sensor models. Empty intervals are caused by sensor readings that do not comply with the given models. This can be due to several factors:

- Systematic or transient sensor errors.
- Inaccurate phenomenon or sensor models. Furthermore, the defined bounds are satisfied with a certain high probability and, hence, also the measurements are invalid within this probability.
- Poor calibration of the sensor measurements.

Assume that an empty interval is received when calculating the intersection between $[c(t_j)]$ and $[c(t_i)]$. Without additional information, we do not know whether the erroneous measurement is $\widehat{m}(t_j)$ or $\widehat{m}(t_i)$. Hence, we assume that only a small portion of the measurements are flawed. Therefore, we build a conflict graph $G = (V, E)$ where two measurements are added to the set of vertices V with an edge in E if the intersection of their signal ranges results in an empty interval. We compute the maximum independent set I of G and, due to our assumption, presume that set I contains the valid measurements.³ The erroneous measurements $[c(t_j)] \in V \setminus I$ are marked as invalid and are not considered in future iterations.

Example. We show in Figure 4.4 an illustrative example with four sensor readings. The measurements are indicated with their signal ranges. We assume a linear phenomenon function f_ω as depicted with dotted lines. On line 25 of the algorithm, we will receive empty intervals because signal range $[c(t_1)]$ neither has an intersection with $[c(t_0)]$ nor with $[c(t_3)]$, e.g.,

³Finding a maximum independent set is NP-hard. We use *Greedy*, which achieves an approximation ratio of $(\Delta + 2)/3$, where Δ is the maximum degree of G [HR97].

$[c(t_0)] \wedge ([c(t_1)] + \Omega_{1,0})$ results in an empty interval. Hence, the algorithm builds a conflict graph, as shown in Figure 4.4(b). Measurements $[c(t_0)]$ and $[c(t_3)]$ build the maximum independent set, therefore measurement $[c(t_1)]$ is marked as invalid.

4.3.3 Convergence

In the following, we show that the algorithm proposed is able to compute for every sensor reading a fixed and bounded sensor accuracy $[\delta(t_j)] \in D$. Similarly, the same can be shown for all signal ranges $[c(t_j)] \in C$.

Theorem 2 (Convergence of the sensor accuracy). *Let S and R be two non-empty sets as given as input in Alg. 1. For every sensor accuracy $[\delta(t_j)] \in D$ the sequence of sensor accuracy lengths $\{\overline{[\delta(t_j)]}^{(n)}\}$ converges to $\inf_n \{\overline{[\delta(t_j)]}^{(n)}\}$.*

Proof. The sequence $\{\overline{[\delta(t_j)]}^{(n)}\}$ is monotonically decreasing, i.e., for all indices n is $\overline{[\delta(t_j)]}^{(n)} \geq \overline{[\delta(t_j)]}^{(n+1)}$, since the only operation on $[\delta(t_j)]$ is the intersection with another sensor accuracy $[\delta(t_i)] \in D$. For the intersection of any given intervals $[a]$, $[b]$, and $[c]$, such that $[c] = [a] \wedge [b]$, the following holds: $\overline{[c]} \leq \overline{[a]}$ and $\overline{[c]} \leq \overline{[b]}$.

Next, we show that sequence $\{\overline{[\delta(t_j)]}^{(n)}\}$ is lower and upper bounded by

$$0 \leq \{\overline{[\delta(t_j)]}^{(n)}\} \leq \overline{[a_r]} + \overline{[\Gamma_{i,j}]} \text{ with } r(t_i) \in R. \quad (4.13)$$

The lower bound 0 is given by the minimum possible size of an interval. The upper bound is narrowed by the (time-wise) closest reference measurement $r(t_i)$ that belongs to a reference pair in R . The initially unbounded interval $[\delta(t_j)]^{(0)}$ is bounded after the $|i - j|$ th iteration to $[\delta(t_j)]^{(i-j)} = r(t_i) + [a_r] + [\Gamma_{i,j}]$ where $[\Gamma_{i,j}]$ represents the maximum possible sensor drift between times t_j and t_i .

After the *Monotone Convergence Theorem*, a monotonically decreasing and bounded sequence $\{\overline{[\delta(t_j)]}^{(n)}\}$ always converges to $\inf_n \{\overline{[\delta(t_j)]}^{(n)}\}$ [Fit06].

□

4.4 Case Study

We present a case study to show the applicability of our signal range algorithm. For this we feed the algorithm with static and mobile temperature and ozone measurements from our installations described in Chapter 2. Temperature is an easily measurable signal whereas the sensing of the ozone concentration is more complex. The gas sensor used is cross-sensitive to other gases in the atmosphere and its sensor readings vary with changing environmental conditions (e.g., temperature).

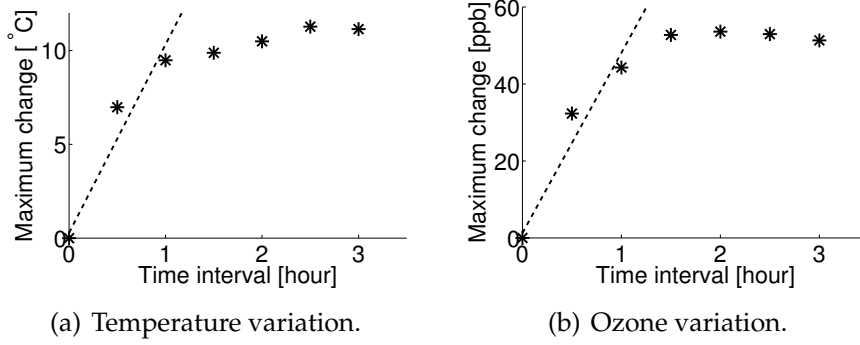


Figure 4.5: Phenomenon gradient. Maximum phenomenon change for time intervals between 30 min and 3 h based on historical high-quality data.

4.4.1 Model Instantiations

We introduce specific instances for the generic models presented in Section 4.2, which are used for the evaluation of our measurements.

Phenomenon model. We introduce the variables ω_t and ω_s to model the maximum feasible phenomenon change from time $t_i \in T$ and location $l_i \in L$ to any other time $t_j \in T$ and location $l_j \in L$. In our data sets, the time intervals between subsequent measurements are in the order of minutes. The evaluation of historical high-quality data of the last 18 months shows, that for short time intervals $|t_i - t_j|$, there is a close to linear dependence between the maximum phenomenon change and the time difference $|t_i - t_j|$, as depicted for temperature and ozone in Figure 4.5 [Nat14b]. Further, we assume that the maximum phenomenon change over a given distance is also linearly dependent on the location offset $|l_i - l_j|$:

$$f_\omega(|t_i - t_j|, |l_i - l_j|) = |t_i - t_j| \cdot \omega_t + |l_i - l_j| \cdot \omega_s. \quad (4.14)$$

An example of such a function f_ω is depicted in Figure 4.2 for a static location. Such linear functions overestimate the maximum possible phenomenon change, especially for large time and location differences, as shown in Figure 4.5. A more precise phenomenon model would further improve the accuracy of the computed signal ranges.

Sensor model. Low-cost sensors are subject to drift over time, which affects their measurement accuracy. We introduce γ_t to model the maximum sensor drift between two measurement times. The maximum accumulated sensor drift from time t_i to t_j is linearly bounded by:

$$f_\gamma(|t_i - t_j|) = |t_i - t_j| \cdot \gamma_t. \quad (4.15)$$

Our signal range algorithm is independent of the phenomenon and sensor models. For the evaluation we use the simple models presented

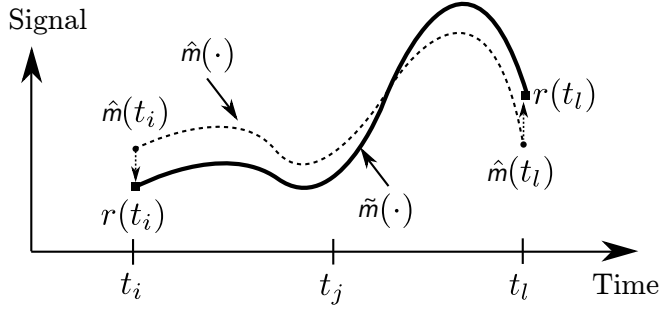


Figure 4.6: Offset adjustment. The reference measurements $r(t_i)$ and $r(t_l)$ linearly adjust the offset of all intermediate sensor readings $\widehat{m}(\cdot)$ resulting in offset adjusted sensor readings $\widetilde{m}(\cdot)$.

above. However, also arbitrary complex models, such as [SW02], can be used to lower the interval bounds of $[\Omega_{i,j}]$ and $[\Gamma_{i,j}]$, and hence, get tighter signal ranges without any changes to the algorithm described in Alg. 1.

Offset adjustment. The proposed signal range algorithm is independent of a specific sensor calibration approach. We increase the accuracy of the derived signal ranges by using the available reference pairs $(r(t_i), \widehat{m}(t_i))$ to adjust the offset of the calibrated measurements. We denote the adjusted measurements with $\widetilde{m}(t_i)$. We use the reference measurements to adjust the offset of the sensor readings, as illustrated in Figure 4.6. We consider two consecutive measurement pairs at times t_i and t_l , and set the calibrated sensor readings equal to the reference signals, i.e., $\widetilde{m}(t_i) = r(t_i)$ and $\widetilde{m}(t_l) = r(t_l)$. The offset adjustment of sensor reading $\widehat{m}(t_j)$, where $t_i < t_j < t_l$, is influenced by the immediate preceding and succeeding reference measurements $r(t_i)$ and $r(t_l)$. The exact factor is linearly dependent on the absolute time differences $|t_j - t_i|$ and $|t_j - t_l|$:

$$\begin{aligned} \widetilde{m}(t_j) = & \widehat{m}(t_j) + \frac{|t_j - t_i|}{|t_l - t_i|} \cdot (r(t_l) - \widehat{m}(t_l)) + \\ & \frac{|t_j - t_l|}{|t_l - t_i|} \cdot (r(t_i) - \widehat{m}(t_i)), \quad \forall t_i < t_j < t_l \end{aligned} \quad (4.16)$$

For example, if measurement $\widehat{m}(t_j)$ is performed exactly between t_i and t_l , then both reference measurements $r(t_i)$ and $r(t_l)$ have 50% influence on the offset adjustment of measurement $\widetilde{m}(t_j)$.

4.5 Evaluation

First, we show the feasibility of the proposed signal range algorithm under a controlled static setting. We thoroughly analyze the accuracy of

the computed signal ranges and the ability to detect sensor failures. Then, we use the found parameter settings to calculate the signal ranges in a mobile monitoring scenario.

Performance metrics. We evaluate the performance of our algorithm by looking at two opposing metrics: (i) the average signal range size and (ii) the ratio of invalid measurements. Assume set M contains all valid signal ranges and set N all invalid ones. We define the average signal range size as

$$\frac{1}{|M|} \sum_{\forall [c(t_j)] \in M} \overline{[c(t_j)]}. \quad (4.17)$$

The average signal range size is highly dependent on the model parameters ω_t , ω_s , γ_t , and e_n . Setting these parameters tight results in smaller signal ranges. However, this also potentially results in a larger number of invalid sensor readings that do not comply with the given models. We define the ratio of invalid measurements as

$$\frac{|N|}{|M| + |N|}. \quad (4.18)$$

Guidance on parameter setting. The proper setting of the model parameters is crucial to derive exact accuracy bounds and accurately detect outliers. The sensor drift over time γ_t and the expected noise in the sensor readings e_n are usually specified in the data sheets of sensors, e.g., [Sen14]. If available, comparative measurements against a higher quality sensor can be used to verify and refine the given specifications. Public historical high-quality data (e.g., from a national air pollution monitoring network [Nat14b]) can be used to compute the parameters ω_t and ω_s that describe the maximum feasible phenomenon change over a given temporal and spatial distance. In the following, we show how the parameter settings of the sensor and phenomenon models impact the performance of the introduced signal range algorithm. Additionally, we present for each data set detailed results for one specific parametrization, based on sensor characteristics given in the data sheets and phenomenon variations found in historical high-quality data.

4.5.1 Static Setting

Measurement setup. We use measurements from our static sensor testbed introduced in Chapter 2. We assume that the fixed station has near perfect sensors, i.e., $[a_r] = [0, 0]$. It measures the temperature and ozone concentrations once per minute and persistently stores 10-minute averages. We use one measurement per hour as input for the signal range

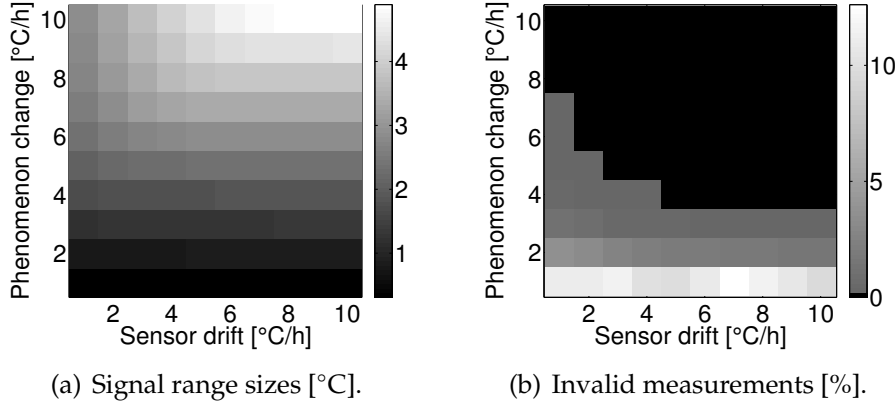


Figure 4.7: Temperature. Average signal range sizes (a) and ratios of invalid measurements (b) for different model settings.

algorithm and the remaining five measurements per hour to evaluate the accuracy of the calculated signal ranges. Our sensor node also monitors the ambient temperature and ozone concentration once per minute. We analyze data collected over a period of 1.5 years comprising 1.5 million sensor readings to reveal seasonal characteristics of the deployed sensors. The detailed results are based on a one-month extract thereof.

Our evaluation metrics depend on the given model parameter settings. We evaluate both metrics with various parameterizations. We analyze different phenomenon models by varying ω_t , the maximum feasible phenomenon change over time. Parameter ω_s has no influence since we are performing measurements at a static location. For the sensor model we vary the maximum sensor drift γ_t . We fix the sensor noise to $e_n = 1.5^\circ\text{C}$ [Sen14] and $e_n = 6$ ppb for the models of the temperature and ozone sensors, respectively.⁴

Evaluation: Temperature measurements. We illustrate in Figure 4.7 the average sizes of the signal ranges and the ratios of invalid measurements for ω_t and γ_t between 1°C/h and 10°C/h . Figure 4.7(a) demonstrates that the average signal range sizes distinctly decrease for small parameter values. However, these tight parameter settings are paid by an increased number of invalid sensor readings that do not confirm the strict model settings, as shown in Figure 4.7(b). For relaxed model parameters, the average signal range sizes are larger but the percentage of invalid sensor readings drop towards zero. In the following, we show detailed results for one specific parameter setting. We fix the sensor drift to $\gamma_t = 0.04^\circ\text{C/yr}$ as specified in the sensor’s data sheet [Sen14] and the

⁴The data sheet of the ozone sensor suggest that $e_n = 20$ ppb [SGX14], however our evaluations show that the noise is around ± 6 ppb.

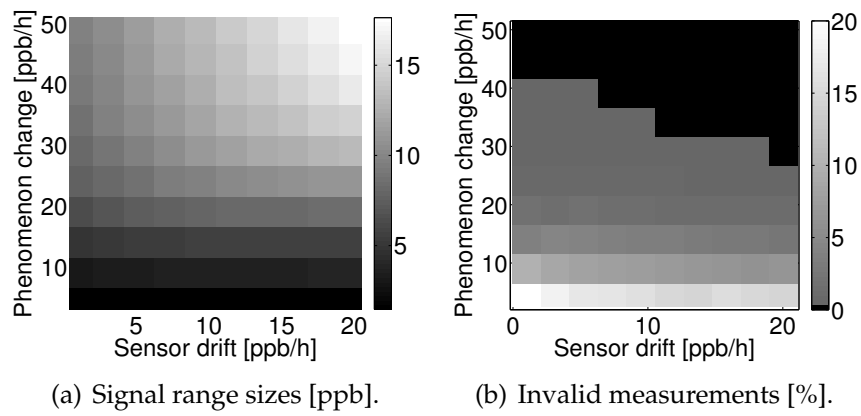
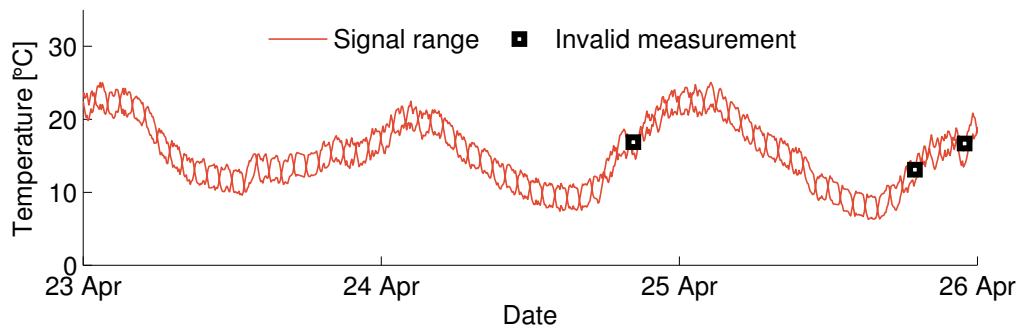
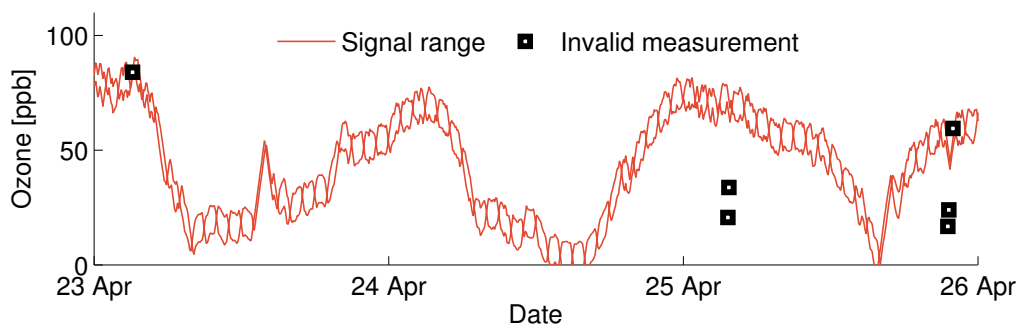


Figure 4.8: Ozone. Average signal range sizes (a) and ratios of invalid measurements (b) for different model settings.

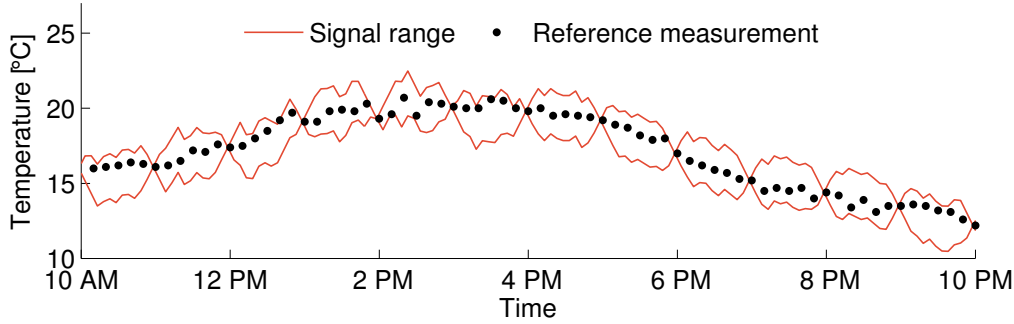


(a) Temperature signal ranges.

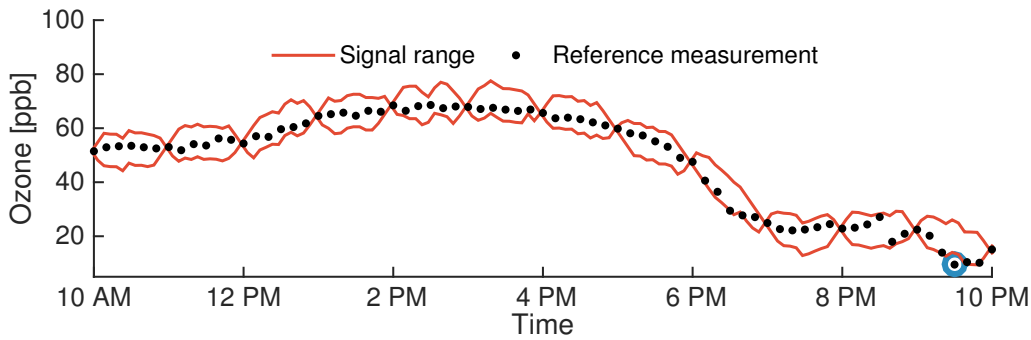


(b) Ozone signal ranges.

Figure 4.9: Static setting. Signal ranges of temperature and ozone measurements over a period of three days. Invalid sensor readings are marked with a square.



(a) Temperature measurements.



(b) Ozone measurements.

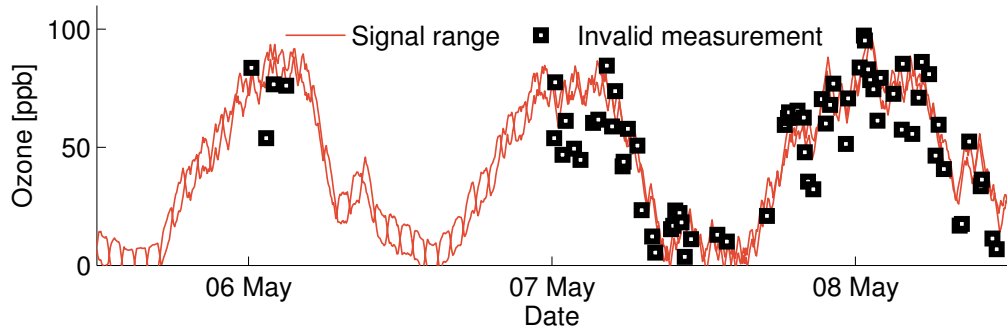
Figure 4.10: Signal range accuracy. The signal ranges of temperature and ozone measurements are accurate, as they closely surround the reference measurements from a static high-quality station.

maximum temperature variation to $\omega_t = 10^\circ\text{C}/\text{h}$ as observed on historical high-quality data (see Figure 4.5(a)). Figure 4.9(a) shows the calculated signal ranges for an excerpt of three days. We get a signal range size of $2.4 \pm 0.9^\circ\text{C}$ and 0.2% of the sensor readings are invalid.

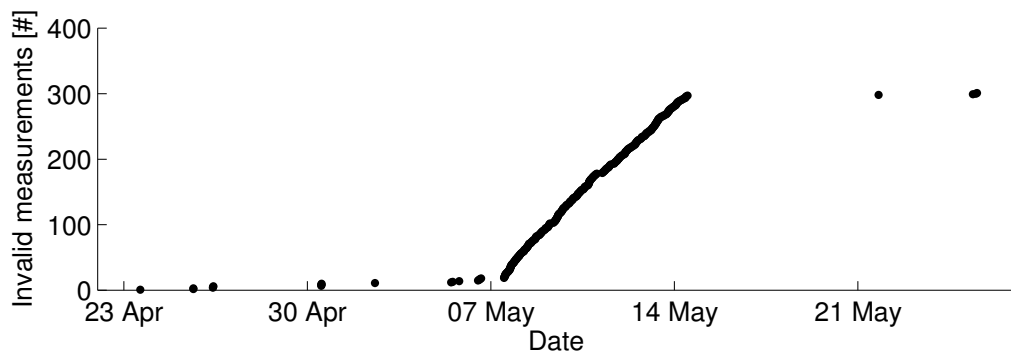
Evaluation: Ozone measurements. We perform the same evaluation on the ozone data set. We depict in Figure 4.8 the average signal range sizes and the ratios of invalid measurements for different parameter settings of ω_t and γ_t . The ratios of invalid measurements are slightly higher than for the temperature measurements. This is not surprising as sensing ozone is more complex and error-prone (e.g., the sensor is sensitive to other gases in the atmosphere).

We set the sensor drift to $\gamma_t = 1\text{ppb}/\text{day}$ ⁵ and the maximum phenomenon variation over time to $\omega_t = 40\text{ppb}/\text{h}$, as observed on historical high-quality data (see Figure 4.5(b)). We receive with this setting an average signal range size of $9.5 \pm 3.7\text{ppb}$ and 0.3% of invalid

⁵Based on the observed resistance decrease of tin dioxide sensors in [HZG11].



(a) Ozone signal ranges and detected invalid measurements (extract).



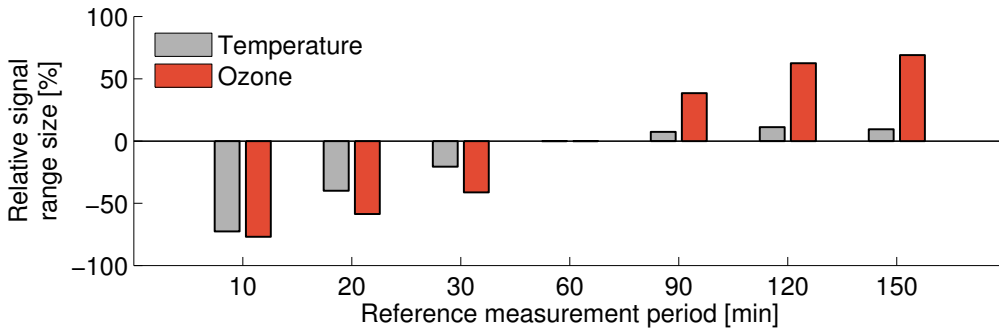
(b) Accumulated number of invalid measurements.

Figure 4.11: Sensor failure detection. The artificially added noise during 7–14 May significantly increases the number of invalid measurements.

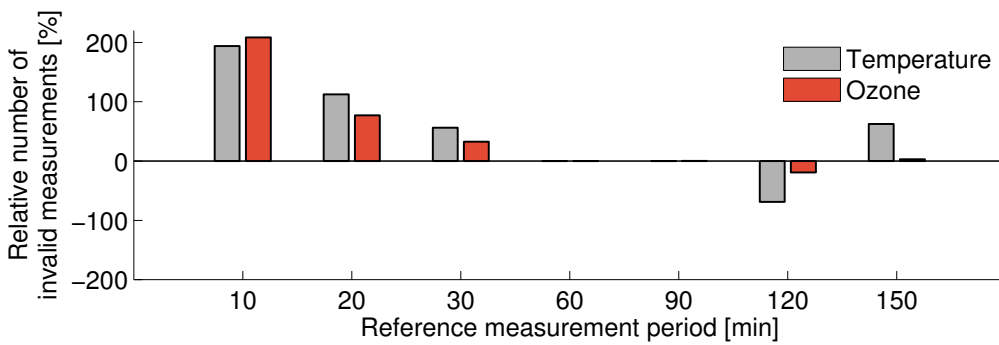
measurements, as illustrated in Figure 4.9(b). We see that clear outliers are correctly marked as invalid, e.g., midday on the 25th of April.

Accuracy. We use all available reference measurements (10-minute averages) to evaluate the accuracy of the calculated signal ranges over the course of one month. Figure 4.10 depicts a half-day excerpt of this data. We clearly see that the signal ranges closely surround the reference measurements. For temperature, 96 % (4542 out of 4739) of the reference measurements are inside the signal ranges. For ozone measurements, 92 % (4238 out of 4608) are inside the computed signal ranges. The percentage is slightly lower as the ozone phenomenon is more complex and hence, our basic models less accurate. However, values outside the signal range miss the bounds only by a short distance, that is on average 0.7°C for temperature and 3.7 ppb for ozone (e.g., the measurement shortly before 10PM in Figure 4.10(b)).

Detecting sensor failures. We demonstrate on partially synthetic data that our method is able to accurately detect and filter out erroneous



(a) Relative signal range sizes.

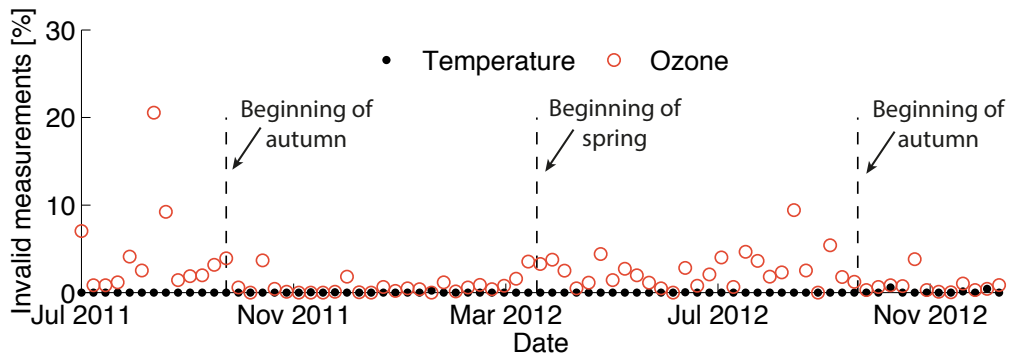


(b) Relative number of invalid measurements.

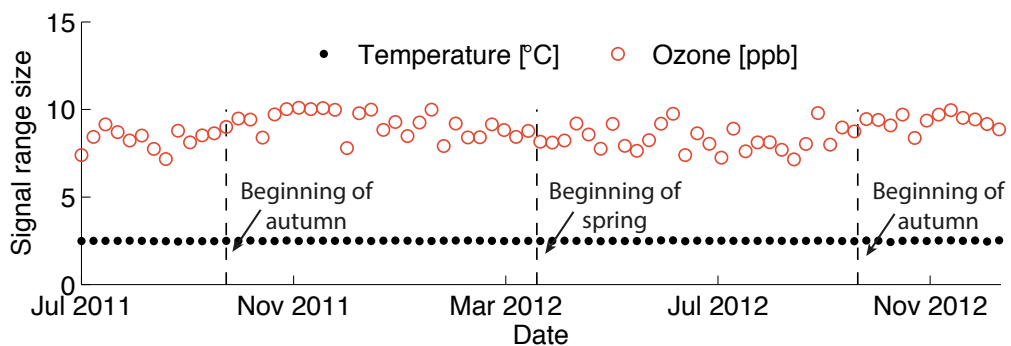
Figure 4.12: Reference data. Relative average signal range sizes and relative number of invalid measurements depending on the number of available reference measurements. Baseline is one reference measurement every 60 min.

measurements and in this way support the detection of sensor failures. We add an offset, which is uniformly distributed in $\mathcal{U}(-20, 20)$ ppb, to all ozone measurements dating from the 7th to the 14th of May. Our method correctly detects the sensor readings that do not comply with the models and marks them invalid, as shown in Figure 4.11(a). Figure 4.11(b) depicts the accumulated number of invalid measurements over time. The steep slope between 7th and 14th of May clearly indicates the period of time with erroneous sensor readings and, hence, can be used as indicator for defective sensors.

Influence of reference measurements. In the following, we analyze the influence of the number of available reference measurements on the computed signal ranges. Figure 4.12 shows the relative average signal range sizes and number of invalid measurements for reference measurement periods between 10 minutes and 2.5 hours. We take as baseline the setting used until now, that is, one reference measurement per hour. More reference measurements allow to calculate more accurate signal ranges, as shown in Figure 4.12(a). The influence on



(a) Weekly ratio of invalid measurements.



(b) Weekly average signal range sizes.

Figure 4.13: Long-term characteristic. The temperature sensor does not show any dependency on seasonal weather conditions while the ozone sensor delivers a higher number of invalid sensor readings in the warm seasons from March to September with higher exposure levels.

the ozone signal ranges is higher than on those of temperature due to the considerably higher drift of the ozone sensor. More reference measurements also help to find up to two times more erroneous sensor readings, as depicted in Figure 4.12(b). The phenomenon model helps to find outliers while due to the reference measurements we are able to also detect many systematic errors (e.g., wrong offset).

Long-term sensor characteristic. We end our analysis by looking at the long-term characteristic of the temperature and ozone sensors. To this end, we evaluate for a period of 1.5 years (from July 2011 to December 2012) for each week the ratio of invalid temperature and ozone sensor readings, as depicted in Figure 4.13(a). The reliability of the temperate sensor does not show any dependency on seasonal weather conditions. In turn, the ozone sensor is more sensitive to environmental changes. It delivers distinctly higher numbers of erroneous measurements during

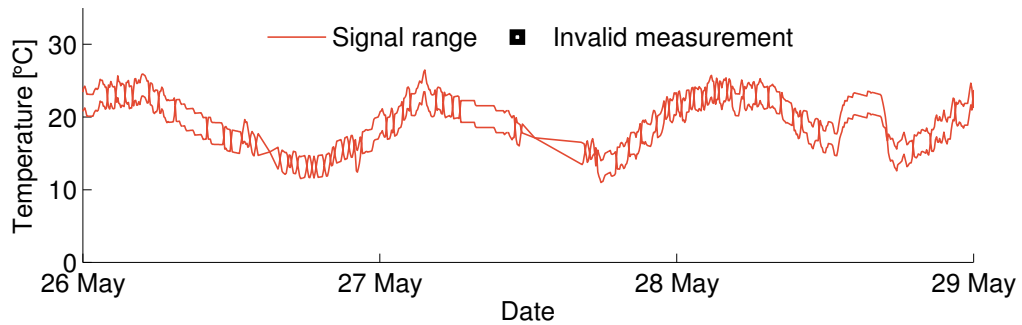
the warm seasons from March to September. This is probably due to the higher ozone levels in summer. The average signal range sizes of the temperature and ozone sensor readings do not reveal any dependency, as shown in Figure 4.13(b).

4.5.2 Mobile Setting

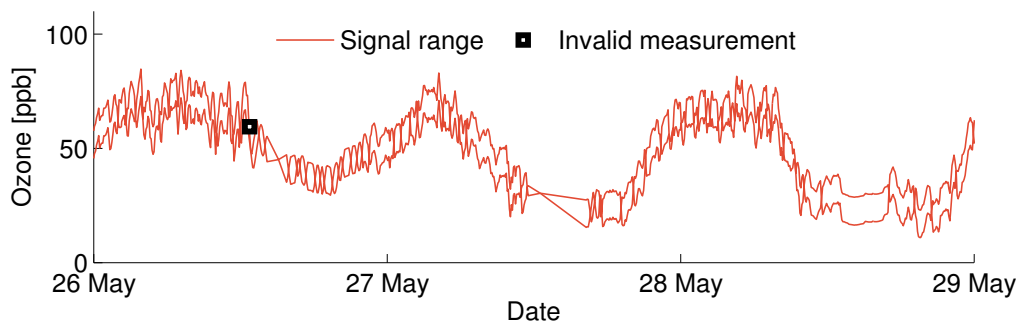
Setup. We analyze measurements from our mobile air pollution monitoring network introduced in Chapter 2. We evaluate four weeks of measurements from five sensor nodes collected in May and June 2012. We use the same parameter settings as in the static analysis and additionally set the phenomenon gradient over space ω_s to $1^\circ\text{C}/\text{km}$ and $9\text{ ppb}/\text{km}$ for temperature and ozone, respectively. Both settings are based on the evaluation of historical data from two static high-quality stations located within 5 km from each other. From time to time the mobile vehicles pass by these two reference stations. We use their measurements as input for the signal range algorithm. We create a reference pair each time the sensor node is in the vicinity of one of the reference stations. The temporal and spatial vicinity requirements heavily depend on the measured phenomenon. The spatial dispersion of temperature and ozone in a street canyon is in general constant and the monitored phenomena are slowly changing over time (in the order of minutes) [VFPGF03]. Hence, we expect that sensor reading and reference measurement can be considered to be exposed to similar concentrations if their measurement time and location do not differ more than 10 min and 400 m.

Evaluation. We illustrate the evaluated signal ranges for temperature and ozone measurements in Figure 4.14. The average signal range sizes are with $2.7\pm 0.9^\circ\text{C}$ for temperature and $10.1\pm 4.2\text{ ppb}$ for ozone, only slightly higher than in the static scenario. In both data sets, 0.3% of the measurements do not comply with the given models.

Accuracy. We perform a 10-fold cross-validation test to evaluate the accuracy of the computed signal ranges. We randomly remove 10% of the reference measurements and use them to verify the accuracy of the calculated signal ranges. We repeat this procedure ten times. For temperature, on average 94% of the reference measurements are inside the computed signal ranges. For ozone, on average 84% of the reference measurements are inside the evaluated ranges. Both percentages are lower than in the static deployment. Possible sources for the lower precision are the sensor response times, which have a much higher influence in the mobile setting, and the vicinity requirements, which may be too wide for certain time periods, leading to inaccurate reference pairs.



(a) Temperature signal ranges.



(b) Ozone signal ranges.

Figure 4.14: Mobile setting. Signal ranges of mobile temperature and ozone measurements of one mobile sensor node over a period of three days.

However, the signal bounds are again missed only by a short distance, that is, on average 0.8°C for temperature and 5.2 ppb for ozone.

4.6 Related Work

Nowadays, wireless sensor networks are used in a wide range of application domains [KWL⁺11, CMP⁺09, CCD⁺11, LLL⁺09, CLBR10]. There has been much research to increase the quality of unstable low-cost sensor readings by employing intelligent calibration algorithms, such as the in-place sensor calibration algorithm from Bychkovskiy et al. [BMEP03] and collaborative sensor calibration from Xiang et al. [XBP⁺12]. However, to the best of our knowledge, we are the first to evaluate the exact quality of individual noisy sensor readings. We propose a method to compute the quality of individual measurement, which can be used for example to evaluate the goodness of a calibration scheme. Cheng and Prabhakar [CP03] and Reznik and Kreinovich [RK04] propose

probabilistic methods to automatically detect sensor readings that reveal uncertainties above a given threshold and improve data accuracy and reliability. We combine phenomenon and sensor models with knowledge on the monitored signal to calculate signal bounds. Also Elnahrawy and Nath [EN03a] combine the noise characteristic of the sensor and prior knowledge of true sensor readings to compute probabilistic confidence bounds. However, they do not incorporate sensor drifts, which can have a significant impact when using low-cost gas sensors.

There is a vast body of literature on outlier detection techniques in general [HA04] and for wireless sensor networks in particular [ZMH10]. Most of the proposed outlier detection algorithms can efficiently handle transient sensor errors, caused by faulty sensor readings that appear abnormal [SGG10]. Different techniques exist to find transient errors, such as done by Elnahrawy and Nath [EN03b] by learning spatio-temporal data correlations, Sheng et al. [SLMJ07] by using histogram-based data abstractions, and Mukhopadhyay et al. [MSPD09] by predicting the next sensor reading based on past measurements. Our approach differs from previous work in the way that we are also able to detect systematic sensor errors (i.e., faulty calibration) using the available reference measurements.

4.7 Conclusions

Knowing the accuracy of collected measurements is of great use for many wireless sensor network applications, where due to limited resources (e.g., energy and bandwidth) low-cost sensors are used that deliver noisy sensor readings and suffer from sensor drift over time. In this chapter, we propose a method to calculate safe accuracy bounds for individual sensor readings. Further, we automatically detect and filter out systematic and transient sensor errors. Our approach is model-driven as it is supported by models that describe the phenomenon of interest and the sensor characteristic in order to compute accuracy bounds with the help of at least one high-quality reference measurement. The models are independent of the signal range algorithm and can be simple or elaborate, as desired. Moreover, the proposed method does not incur any additional overhead on the sensor nodes.

Our evaluation of static and mobile temperature and ozone measurements demonstrates the applicability of our method. The comparison with high-quality reference measurements reveals the good precision of the computed signal ranges. Up to 96 % of the temperature and up to 92 % of the ozone reference measurements are inside the evaluated signal ranges. Further, we revealed how the amount of available reference data

influences the size of the signal ranges and the detection of erroneous sensor readings. By analyzing 1.5 years of temperature and ozone measurements from our sensor testbed we showed that the ozone sensor's reliability depends on seasonal changes of environmental conditions.

5

Developing Fine-Grained Air Pollution Maps

Up-to-date information on urban air pollution is of great importance, e.g., for environmental protection agencies to assess air quality and provide advice to the general public in a timely manner. In this chapter, we focus on the processing of our ultrafine particle (UFP) measurements as they represent a very interesting data set (however, the methods described are applicable to all pollution data sets collected with our deployments). UFPs are widely spread in urban environments and may have a severe impact on human health. The lack of knowledge about the spatio-temporal distribution of UFPs hampers profound evaluation of these effects. We analyze one of the largest spatially resolved UFP data set publicly available today containing over 50 million measurements. We collected the measurements throughout more than two years using our streetcar deployment. Based on these data, we develop land-use regression models to create pollution maps with a high spatial resolution. We compare the accuracy of the derived models across various time scales and observe a rapid drop in accuracy for maps with sub-weekly temporal resolution. To address this problem, we propose a novel modeling approach that incorporates past measurements annotated with metadata into the modeling process. In this way, we achieve a 26% reduction in the root-mean-square error—a standard metric to evaluate the accuracy of air quality models—of pollution maps with semi-daily temporal resolution. We believe that our findings can help epidemiologists to better understand the adverse health effects related to air pollutants and serve as a stepping stone towards detailed real-time pollution assessment.

5.1 Introduction

Most countries have mass emission limits for particulate matter PM_{10} and $PM_{2.5}$ (i.e., particles with a diameter of less than $10\ \mu\text{m}$ and $2.5\ \mu\text{m}$, respectively), but have no restrictions on ultrafine particles (UFPs). UFPs are particles with a diameter of less than $100\ \text{nm}$. In ambient air, UFPs are mainly man-made as byproducts of specific high temperature processes, such as combustion reactions in car engines. The adverse health effects of UFPs are most probably underestimated when they are traditionally monitored by mass as part of PM_{10} and $PM_{2.5}$ [PTT⁺01]. This is because UFPs make a dominant contribution to the total number of urban particle concentrations, but their contribution to the total particle mass is small [HY00]. Therefore, UFPs were not considered particularly hazardous in the past. There are strong indications, however, that adverse health effects are more related to particle number concentration rather than to particle mass [PTT⁺01]. To better understand the adverse health effects of UFPs, it is essential to have spatially resolved UFP concentration measurements at hand [HBK⁺09].

Nowadays, air pollution is monitored by networks of static measurement stations operated by official authorities. These stations are highly reliable and able to accurately measure a wide range of air pollutants. However, their high acquisition and maintenance costs severely limit the number of installations. As a result, very little is known about the spatial distribution of air pollutants in urban environments and there is a lack of accurate intraurban air pollution maps. However, for air pollutants with high spatial variability, such as UFPs, the public availability of reliable pollution maps is essential. They raise the citizens' awareness about air pollution, empower environmental scientists to craft and evaluate new policies, and enable in combination with other spatial data sets new kind of applications (see Chapter 6).

Contributions and road-map. To tackle the challenges above, we use measurements from our mobile streetcar monitoring network described in Chapter 2. Node mobility trades off temporal resolution against spatial resolution, enabling a high spatial resolution across large areas without the need for a huge number of fixed sensors. However, due to the reduced temporal resolution of any covered location, it is a formidable challenge to derive pollution maps with a high temporal resolution at daily or hourly time scales. In this chapter, we demonstrate that a mobile measurement system can effectively be used to derive accurate pollution maps with a high spatial and temporal resolution.

Our mobile measurement system consists of ten sensor nodes installed on top of streetcars of the public transport company in Zurich,

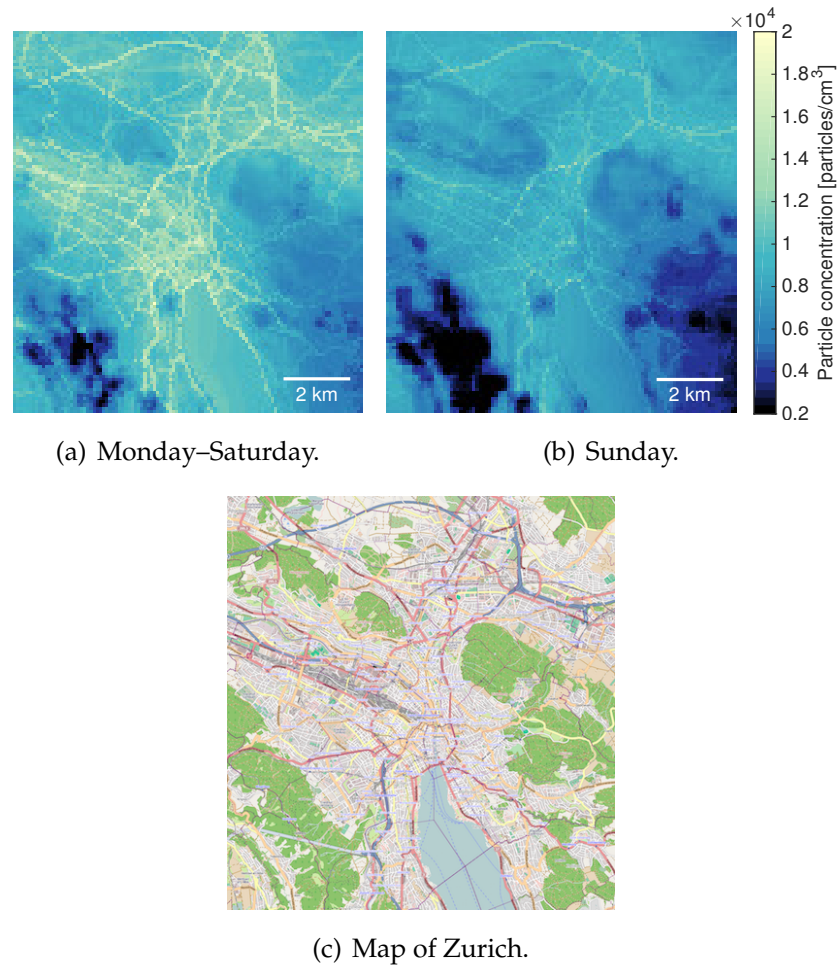


Figure 5.1: Novel ultrafine particle concentration maps for Zurich (Switzerland). The particle concentrations are higher during the week (Monday–Saturday) than on weekends (Sunday) due to higher traffic volumes.

Switzerland. The sensor nodes are equipped with a novel measurement device (MiniDiSCs [FHSB11]) to monitor UFP concentrations. Throughout more than two years, we collected over 50 million UFP measurements. Based on these data, we develop land-use regression (LUR) models to produce accurate pollution maps with high spatio-temporal resolution, such as those depicted in Figure 5.1, where we can clearly spot that, on average, air pollution is higher during the week (Monday–Saturday) than on weekends (Sunday). LUR models use a set of explanatory variables (land-use and traffic data) to model pollution concentrations at locations not covered by the mobile sensor nodes. In a first step, we evaluate the dependencies between the explanatory variables and the measurements. Then, we exploit these relationships to predict the pollution levels for all locations without measurements but with available

land-use and traffic information. Using this method and our mobile measurement system, we derive accurate and fine-grained pollution maps, which are valuable to environmental scientists, epidemiologists, and the general public. We will use the developed pollution maps in Chapter 6 to analyze how much urban dwellers from Zurich (Switzerland) can reduce their exposure to UFPs by not taking the shortest path between origin and destination but a healthier and slightly longer alternative route.

In summary, this chapter makes the following contributions:

- We introduce in Section 5.2 the large set of UFP measurements collected with our mobile measurement system deployed on top of streetcars in the city of Zurich (Switzerland). From April 2012 to April 2014, we collected more than 50 million UFP measurements.
- Assessing the quality of the measurements is difficult due to very sparse ground truth data. We post-process the measurements (calibration and filtering) and propose in Section 5.3 a three-fold validation approach to evaluate the quality of the processed data. Our analysis indicates a high data quality.
- We use the validated measurements in Section 5.4 to derive LUR models for UFP pollution maps with a high spatial resolution of $100\text{ m} \times 100\text{ m}$. In Section 5.5 we apply standard metrics to analyze the quality of the models from yearly up to semi-daily temporal resolutions. We find a good quality of pollution maps with yearly to weekly time scales, while models with sub-weekly temporal resolutions perform less well.
- To tackle this problem, we propose in Section 5.6 a novel modeling approach that incorporates past measurements (annotated with metadata, such as environmental and meteorological conditions) into the modeling process. In this way, we increase the quality of pollution maps with a high temporal resolution. For example, we decrease the root-mean-square error—a standard metric to evaluate the accuracy of air quality models—of semi-daily pollution maps by 26%.

Using our measurement system and modeling approach, we create UFP pollution maps with an accuracy that is comparable to state-of-the-art air pollution maps, while achieving unprecedented spatio-temporal resolution [HBdH⁺08]. We survey related work in Section 5.7, and conclude in Section 5.8.

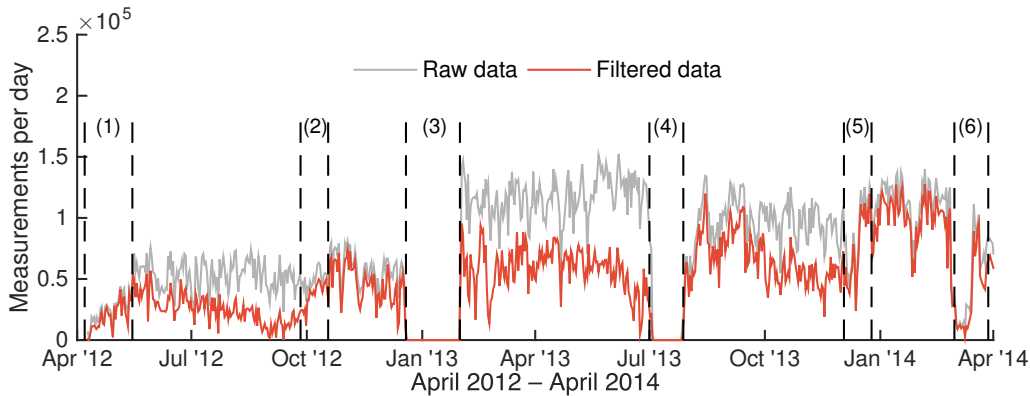
5.2 Spatially Highly Resolved Ultrafine Particle (UFP) Measurements

We analyze a large set of ultrafine particle (UFP) measurements collected with our mobile sensor nodes installed on top of public transport vehicles in Zurich, Switzerland (see description in Section 2.3.1). The sensor nodes deployed are equipped with Miniature Diffusion Size Classifiers (MiniDiSCs) [FHSB11] (see Figure 2.4(a)), a novel tool for UFP monitoring. Traditionally, airborne particles are monitored by mass, hence UFPs, which appear in high numbers but have low mass, are not well represented, despite their specific adverse health effects [PTT⁺01]. Whereas the MiniDiSC, a compact hand-held device (4 x 9 x 18 cm), is able to monitor *particle number concentrations*. It can detect concentrations between 10^3 – 10^6 particles/cm³, so it embraces the average daily range in urban environments of 10^4 – 10^5 particles/cm³ [ZHKS02, RDSF⁺14].

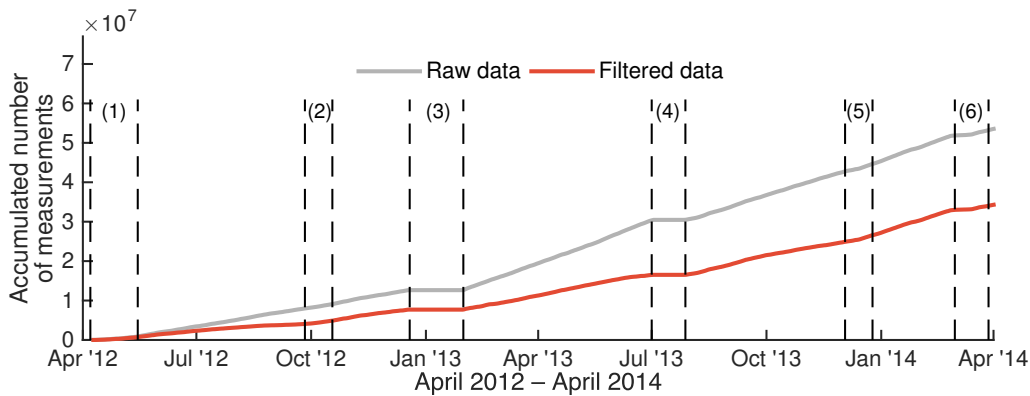
The functional principle of the MiniDiSC is based on unipolar charging of aerosol particles, followed by detection in two electrometer stages. In short, the particles are first charged in a standard positive unipolar diffusion charger, which imparts an average charge on the particles that is approximately proportional to the particle diameter. The charged particles then flow through a diffusion stage and generate a current, which is dependent on the average particle size and is used to calculate the number of particles per cm³.

5.2.1 Data Set

We deployed the first five sensor nodes with integrated MiniDiSCs in April and May 2012 and additional five devices in January 2013. The MiniDiSCs sample UFP every 50 ms. The measurements are aggregated to one sample per 5 s, to reduce the amount of transmitted data. Each sensor node transmits around 10,000 measurements per day to the back-end infrastructure, as depicted in Figure 5.2(a) (raw data). In total, we collected over 50 million aggregated measurements over a time period of two years, as shown in Figure 5.2(b). Further, we depict in Figure 5.2 the six main service phases since the start of the deployment: (1) Initial installation of the first five sensor nodes, (2) cleaning of the MiniDiSCs on top of the streetcars (no deinstallation required) to increase data quality, (3) deinstallation of the MiniDiSCs for cleaning and re-calibration, and installation of five additional devices, (4) deinstallation of all devices for cleaning and recalibration, (5) cleaning of the MiniDiSCs on top of the streetcars, and (6) sensor upgrade with new carbon monoxide (CO) and nitrogen dioxide (NO₂) sensors.



(a) Number of measurements per day across all installed devices.



(b) Accumulated number of measurements across all installed devices.

Figure 5.2: Number of UFP measurements per day and accumulated number of measurements over the course of two years with six service phases (1)–(6).

5.2.2 Offset Calibration and Filtering of the Data

We ensure a high data quality by calibrating and filtering the timestamped and geo-tagged measurements. For one minute in every hour the devices go into a self-calibration phase to measure their null-offsets, which we use offline to adjust the offset of all measured particle concentrations. Then, we use a two-stage filtering process to remove faulty and unreliable measurements. Note that periodic calibration, as described in Chapter 3, is not required for the UFP measurement device (and would also not be possible as the single fixed station monitoring UFPs in Zurich is not located close enough to the streetcar tracks, see Figure 5.3).

First, a GPS-based filter eliminates measurements with horizontal dilution of precision (HDOP) values above 3. The HDOP value specifies the GPS location's precision based on the geometric positioning of the GPS satellites [Lan99]. Values below 3 denote a good to excellent positioning within a few meters. We consider only these measurements, which make up more than 99% of all measurements.

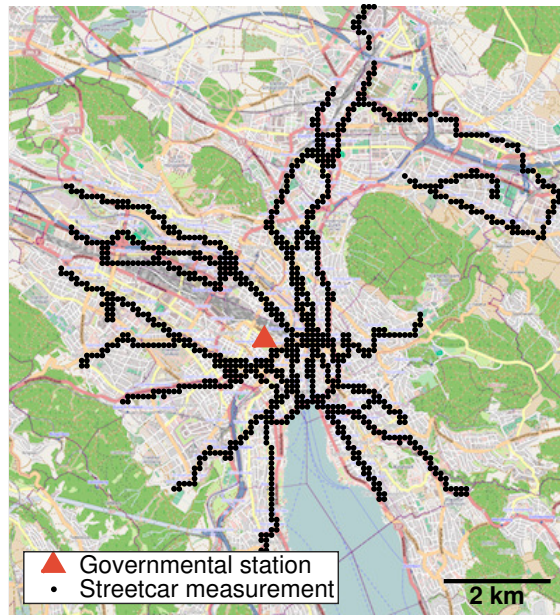


Figure 5.3: Ten mobile sensor nodes deployed on top of public transport vehicles achieve a good coverage of the city of Zurich (Switzerland). The dots denote locations with at least 50 measurements over the course of two years.

The second filter examines the internal status variables of the MiniDiSCs, which are transmitted to the back-end server in addition to the particle concentration numbers. These variables allow to draw conclusions about the proper functioning of the devices. We discard measurements if a status variable indicates a malfunction, such as a too low air flow. Additionally, the MiniDiSCs need a warm-up phase; therefore, all measurements within one hour after start-up are discarded (the nodes are continuously powered 20 h per day on average, as depicted in Figure 2.9). The two-stage filtering process invalidates 36% of the measurements (mainly the second filter), as shown in Figure 5.2. The filtering steps are crucial to achieve high data quality, as we will detail in the next section.

5.2.3 Spatial coverage

A good spatial coverage of the measurements is essential to precisely assess the UFP distribution in urban environments. The spatial coverage of the filtered UFP measurements is shown in Figure 5.3. The dots denote locations (100 m \times 100 m) with at least 50 measurements over the course of two years. The ten mobile sensor nodes achieve a good coverage, in particular compared to the single governmental station monitoring UFPs in Zurich, denoted with a triangle in Figure 5.3. Our measurements cover a large set of diverse location characteristics. For example, the data set

includes measurement locations at terrain elevations from 400–610 m and at diverse traffic densities ranging from vehicle-free zones to areas with over 90,000 vehicles per day.

5.3 Data Validation

Good data quality is a must for the development of reliable pollution maps. Evaluating the quality of measurements spread over a large urban area is a challenging task, especially if no (or only sparse) ground truth is available. The MiniDiSC has been thoroughly tested in laboratory environment and in the field [FHSB11]. However, despite these tests it is unclear whether the harsh deployment setting on top of streetcars has an impact on the quality of the measurements. Among others, the devices have to endure mobility, constant vibrations, high temperature and humidity variations, and long, unattended operating times.

Thus, we propose a three-fold validation approach to assess the quality of our measurements. We *(i)* analyze the statistical distribution of the monitored particle concentrations, *(ii)* evaluate the baseline signal of each device, and *(iii)* compare our measurements to data from two high-quality stations collected during the same time period but at different locations in Switzerland.

5.3.1 Statistical Distribution

The statistical distribution of data from many different scientific disciplines, including the concentration of ambient air pollutants, closely follows a log-normal distribution, i.e., the logarithm of the measurements is normally distributed [LSA01]. Previous work indicates that also UFPs are approximately log-normally distributed [MHCH12]. Here, we confirm that UFPs are log-normally distributed in urban environments. We show in Figure 5.4 the distribution of the raw and processed (i.e., applied calibration and filtering) data and the log-normal distributions with means and standard deviations of the raw and processed data, respectively. Raw data are not log-normally distributed, as apparent from Figure 5.4(a). However, as we show in Figure 5.4(b), the distribution of the processed data nicely fits the log-normal distribution $\text{Log-}\mathcal{N}(\mu, \sigma^2)$ with $\mu = 9.34$ and $\sigma^2 = 0.50$.

5.3.2 Baseline Signal

We examine the correct offset of the MiniDiSCs, by looking at their baseline signals, i.e., low-pass filtered measurements. We expect a similar

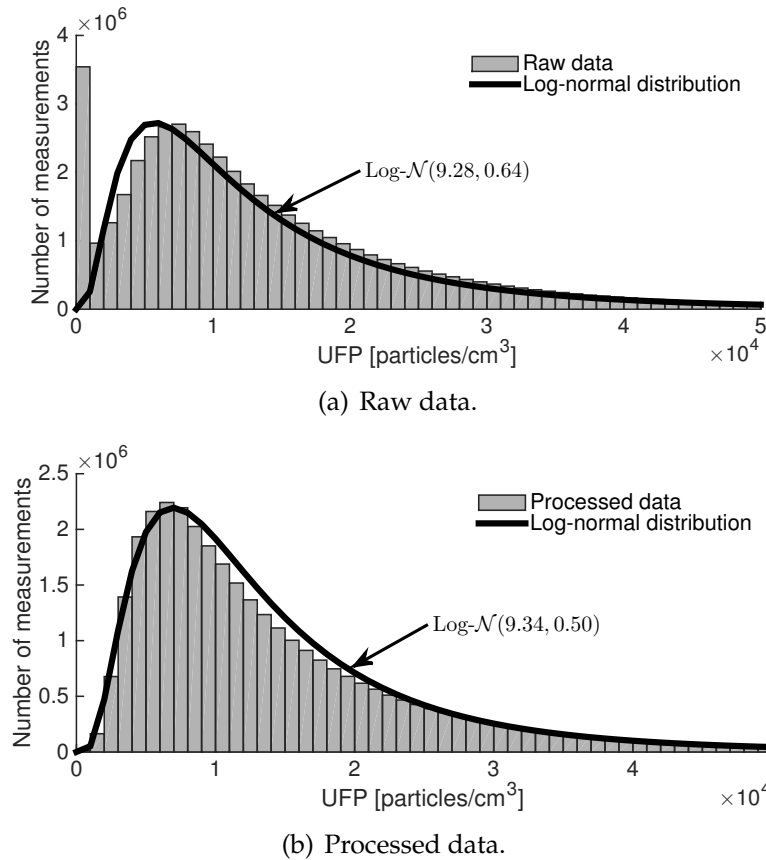


Figure 5.4: The log-normal distribution (black) with mean and standard deviation of the UFP data (gray). After calibration and filtering (processed data), the log-normal distribution accurately fits the distribution of the measurements.

baseline signal across all devices, because all devices take measurements in the same region (streetcars are not bounded to specific lines). We compute the baseline signal of each device by using a low-pass filter. We take for every 90 min time interval, the average time for a streetcar to cross the city twice, the 20th percentile of the measured concentrations. Our analysis confirms that during the two years analyzed the baseline signals are similar across all ten devices.

5.3.3 Comparison to High-Quality Data Sets

We compare our data set to UFP concentrations measured by fixed stations of the Swiss National Air Pollution Monitoring Network (NABEL). Five out of the 16 NABEL stations are equipped with accurate UFP counters (condensation particle counters, CPCs). The stations are located in urban, suburban, and rural environments. For the comparison, we choose data from the stations with urban heavy traffic (Bern–Bollwerk) and suburban (Basel–Binningen) backgrounds, as these locations best

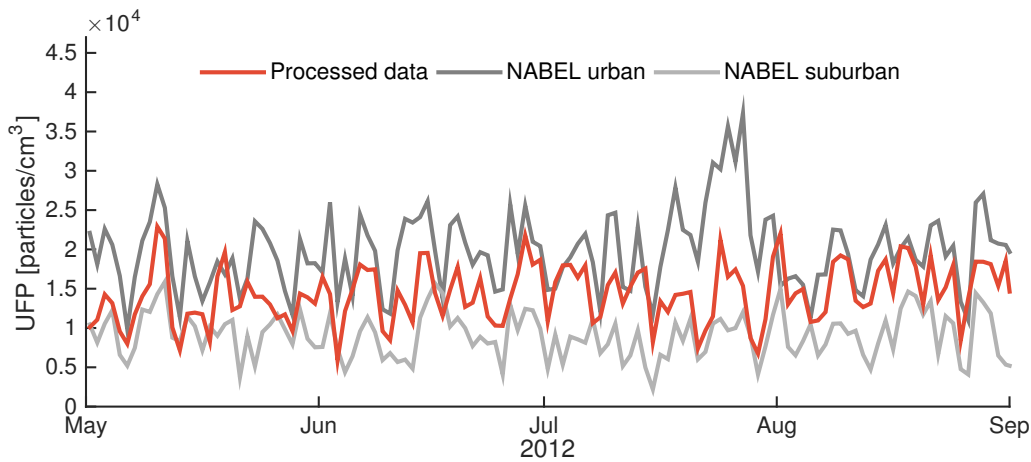


Figure 5.5: Daily average UFP concentration measured by the mobile sensor nodes in Zurich corresponds well to measurements of two fixed NABEL stations at urban (Pearson $r = 0.49$) and suburban ($r = 0.55$) locations.

reflect the topographical extremes of our measurement region in Zurich.

We compare the daily measured average particle concentrations from all mobile sensor nodes to daily averages measured by the two fixed stations. Thereof we depict in Figure 5.5 a segment of four months. It shows that the three data sets have a similar trend, further reflected by the good Pearson correlations of $r = 0.49$ and $r = 0.55$ between our processed data and the urban and suburban located stations, respectively. The two NABEL data sets have a slightly lower correlation of $r = 0.46$. Furthermore, as we would expect, the daily average UFP concentration in Zurich ranges between the daily averages measured at the urban location with heavy traffic and at the suburban location.

The above results, namely the correct statistical distribution, the matching of all devices' baseline signals, and the good correlation and expected range of our data in comparison with the NABEL measurements, indicate a good quality of the processed data. In the following, we use this data set to derive land-use regression models for accurate high-resolution pollution maps.

5.4 Developing Land-Use Regression Models to Create High-Resolution Pollution Maps

Land-use regression (LUR) models are widely used to assess spatial variation of air pollutants, typically at an intraurban scale [JAK⁺04, HBdH⁺08, RDSF⁺14, ZGT⁺14]. LUR models use land-use and traffic

characteristics (explanatory variables) to predict pollution levels for locations not covered by measurement devices. The general concept is based on two steps:

1. At all measurement locations the dependencies between explanatory variables (e.g., population density, traffic volume, and terrain elevation) and monitored pollution levels are evaluated using linear regression.
2. The relationships found between concentrations monitored and the explanatory variables are used to predict concentration levels at locations without measurements but with available land-use data.

There are different approaches to construct LUR models. We use Generalized Additive Models (GAMs) [HT90], because they support non-linear relationships between monitored concentration levels and explanatory variables. Furthermore, GAMs have been used recently to analyze and model the spatio-temporal variability of particulate matter [BHK⁺11, LPK09].

5.4.1 Generalized Additive Models

We divide our data into different time periods and build for each time period a separate model. This yields 989 models with yearly, seasonal, monthly, biweekly, weekly, daily, and semi-daily (midnight–noon and noon–midnight) temporal resolutions. For all models we use the following relationship between pollution concentration and explanatory variables:

$$\ln(c_{\text{num}}) = a + s_1(A_1) + s_2(A_2) + \dots + s_n(A_n) + \epsilon, \quad (5.1)$$

where c_{num} denotes the UFP concentration, a the intercept, $s_1(A_1) \dots s_n(A_n)$ the smooth functions $s_1 \dots s_n$ with explanatory variables $A_1 \dots A_n$, and ϵ the error term. The non-parametric functions $s_1 \dots s_n$ are smooth regression splines. Through empirical evaluations we found that splines with an upper limit of three on the degrees of freedom and the logarithmic link function, $\ln(c_{\text{num}})$ in (5.1), yields the best model residuals, which are the differences between measured and predicted values.

5.4.2 Selecting Explanatory Variables

We examine a set of 12 explanatory variables (listed in Table 5.1), which are typically used to derive LUR models [HBdH⁺08]. Data to calculate population and industry densities, building heights, heating type, and elevation, slope, and aspect of the terrain are from the Swiss Federal

Variable [unit]	Variable [unit]
Population [inhabitants/ha]	Industry [industry buildings/ha]
Building height [floor levels/ha]	Heating [oil and gas heatings/ha]
Terrain elevation [average m/ha]	Road type [busiest road type/ha]*
Distance to next road [m]	Distance to next large road [m] [†]
Terrain slope [average degree/ha]	Terrain aspect [average degree/ha]
Traffic volume [vehicles per day/ha]	Distance to next traffic signal [m]

*Five road types: residential, tertiary, secondary, primary, and freeway.

[†]Road types classified as large: secondary, primary, and freeway.

Table 5.1: These 12 explanatory variables are examined to build the air quality models for UFP pollution maps.

Statistical Office. Road types and distances to next road and traffic signal are extracted from OpenStreetMap data. Average daily traffic volumes are obtained from the Department of Waste, Water, Energy, and Air of the Canton Zurich. Some of the data sets are based on measurements from 2007–2010. We assume that these explanatory variables did not change considerably compared to 2012–2014, which is the monitored and modeled time period. Most explanatory variables have a 100 m x 100 m (1 ha) resolution, which we use for all explanatory variables, as indicated in Table 5.1. This automatically sets the spatial resolution of the generated pollution maps.

It is crucial to examine the correlation between each pair of explanatory variables. Removing variables that have high correlation with each other helps to better distinguish individual contributions of different variables to the modeled particle concentration and prevents redundancy in the model [JCCE09]. We do not use population density and number of gas and oil heating households as both have a strong linear relationship with each other and with the number of floor levels (R^2 is larger than 0.6 [BHK⁺11]). Using the distance to the next traffic signal did not improve any of the tested models, and is, therefore, also removed from the modeling process.

The terms on the right side in (5.1) have a multiplicative relationship with the particle concentration c_{num} , i.e., $c_{\text{num}} = e^a \cdot e^{s_1(A_1)} \cdot e^{s_2(A_2)} \dots$. Thus, we can specify the influence of each explanatory variable A_n as a multiplicative influence factor $e^{s_n(A_n)}$. Figure 5.6 depicts typical influence factors observed in our models. Building height and daily traffic volume have an amplifying while terrain elevation has a reducing impact on the particle concentration. The rugs on the x-axis show that a reduced number of measurements is available at the extremes resulting in decreased confidence bands for these regions.

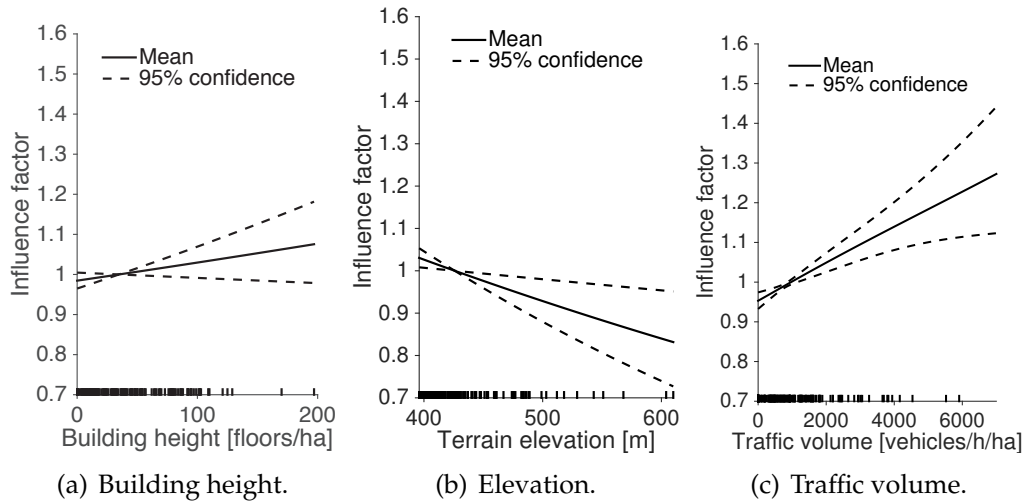


Figure 5.6: Typical influence factors. Building height and traffic volume have a positive while terrain elevation has a negative impact on UFP concentrations.

5.4.3 UFP Data Aggregation

According to the resolution of the explanatory variables, the measured UFP concentrations of two years (April 2012 to March 2014) are projected on a grid with 13,200 cells, each of size $100\text{ m} \times 100\text{ m}$, covering the complete region of interest depicted in Figure 5.3. The subset of the data used depends on the desired temporal resolution of the model, i.e., for daily maps we consider measurements from a single day while for seasonal maps those from the entire season. For each grid cell i , we fetch the n_i measurements located in the cell to calculate the mean pollution concentration c_i^m . The applied GPS filter ensures that the localization is precise enough to correctly assign the measurements to their cells with high probability.

The measurements are (unevenly) distributed among 300–1300 different cells, depending on the analyzed temporal subset and spatial coverage of the data. For example, every black dot in Figure 5.3 denotes a cell with at least 50 measurements when projecting the full two-year data set onto the grid. As model input we use the 200 cells with the highest number of measurements, which are mainly cells containing a streetcar stop. This ensures that the calculated means c_i^m are reliable and provides a good trade-off between spatial input distribution and model performance, as we will show in Section 5.5. Introducing prior weights on the cells (e.g., using number of measurements per cell) does not improve the models.

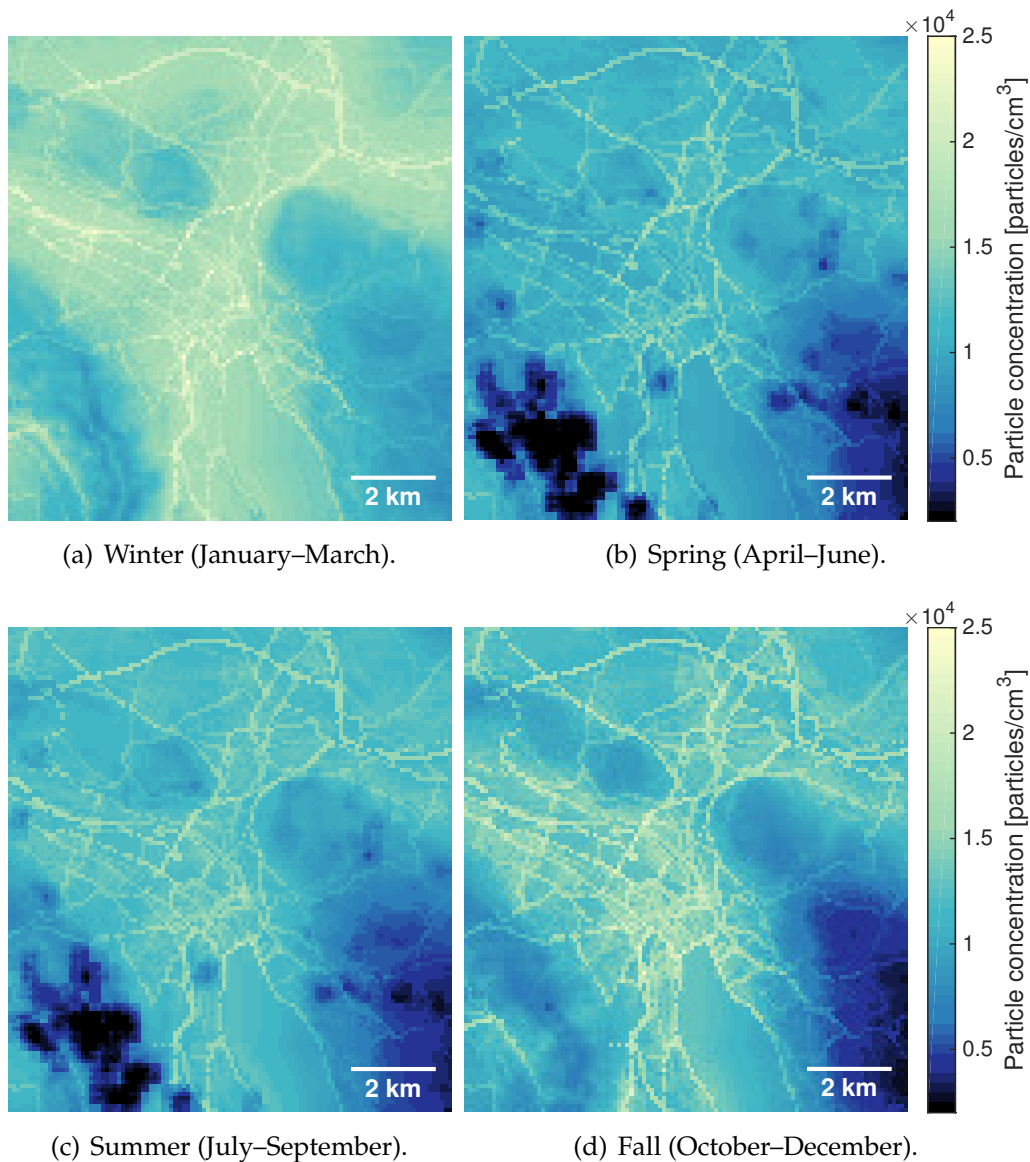


Figure 5.7: Seasonal UFP pollution maps with a spatial resolution of $100\text{ m} \times 100\text{ m}$ for Zurich (Switzerland), based on data collected by mobile sensor nodes throughout a year. Frequent high-inversion fog in winter and fall lead to higher particle concentrations than in spring and summer.

5.4.4 Model Output: High-Resolution Pollution Maps

We use the method described above to develop models for pollution maps with different temporal resolutions. For instance, the output of the models with seasonal resolution (winter, spring, summer, and fall) is shown in Figure 5.7. In general, terrain elevation, building heights, and traffic density have the highest influence on the predicted pollution levels. Further, we see that the pollution levels are higher in winter

and fall than in spring and summer. This is due to frequent high-inversion fog (fog above the surface, which does not completely dissipate) in Zurich from October to March preventing the pollutants to be lifted from the surface and therewith increasing pollution concentration near the ground [Fed09]. The topographical structure around Zurich makes the city prone to inversion fog as it is closely surrounded by hills in the north-east and south-west.

5.5 Revealing the Temporal Resolution Limit

This section evaluates the performance of the 989 generated air quality models with yearly to semi-daily temporal resolutions based on our measurements on top of streetcars from April 2012 to March 2013 (including eight weeks of maintenance work without measurements). These are 1 yearly, 4 seasonal, 11 monthly, 23 biweekly, 44 weekly, 309 daily, and 597 semi-daily models.¹ Our evaluation reveals the following findings:

- Models for pollution maps with yearly to weekly time scales have an accuracy that is comparable to recently published state-of-the-art air quality models.
- Pollution maps with daily and semi-daily temporal resolutions are less accurate due to the reduced number of measurements available to derive the models.

5.5.1 Metrics and Evaluation Methodology

We use three standard metrics to evaluate our models [CH04].

Factor of 2 measure (FAC2): Quantitatively analyzes scatter plots (predicted concentrations plotted against measured concentrations) by evaluating the fraction of data points that lie inside the factor two area, i.e., fraction of data that satisfy

$$0.5 \leq \frac{c_i^p}{c_i^m} \leq 2.0, \quad (5.2)$$

where c_i^p is the model predicted and c_i^m is the average measured concentration of grid cell i . This measure is based on the assumption that an accurate model for pollution maps should have a relative scatter less than a factor of two [CH04].

¹For some days we do not have two semi-daily models because the nodes were only in operation for part of the day.

Root-mean-square error (RMSE): Quantifies the difference between the predicted and measured particle concentrations:

$$RMSE = \left(\frac{1}{N} \cdot \sum_{i=1}^N (c_i^p - c_i^m)^2 \right)^{0.5}, \quad (5.3)$$

where N denotes the number of cells used in the evaluation.

Adjusted coefficient of determination (R^2): Indicates from 0 to 1 how well the predictions fit the measurements ($R^2 = 1$ denotes a perfect fit). R^2 reflects the linear relationship between predicted and measured values, hence, it is insensitive to additive and multiplicative errors. For example, absolute differences between predicted and measured values do not influence the R^2 value.

For all evaluations we perform a 10-fold cross-validation. That is, we randomly remove 10 % of the measurements and create (calibrate) the model using the remaining 90 % of the data. We use the predictions at the locations of the removed 10 % to evaluate the metrics above. We repeat this procedure 40 times to have a good coverage of the complete data set.

5.5.2 Model Performance

We first ensure that the models have no systematic bias and then use the metrics above to evaluate their accuracy.

Trustworthy models should not have any systematic bias in their predictions. In general, the predictions did not show any bias. We show this exemplary in Figure 5.8 on scatter plots of seasonal pollution maps (see Figure 5.7). The dashed lines denote the factor of two area, i.e., slopes of 0.5 and 2, and the optimal 1-to-1 relation, i.e., slope of 1. The linear regression lines (solid) are very close to the 1-to-1 lines for all depicted seasons, indicating the absence of systematic errors.

Figure 5.9 shows the three metrics as box plots for all analyzed temporal time scales. The central mark of each box plot denotes the median, edges show the 25th and 75th percentiles, whiskers cover the most extreme data points without outliers, and the crosses depict the outliers. Models with yearly to weekly temporal resolutions have very high FAC2s, while daily and semi-daily predictions have a considerable number of models with lower FAC2 values, i.e., many predictions are more than a factor of two off. The RMSEs grow with increasing temporal resolution. Yearly to weekly pollution maps have low RMSEs, which are in range of recently published air quality models [HBdH⁺08]. Models with higher temporal resolutions, such as daily and semi-daily pollution maps, have on average 38 % higher RMSEs.

The average R^2 of yearly to monthly maps is 0.38, and slightly decreases for shorter time scales. Recently published air quality models

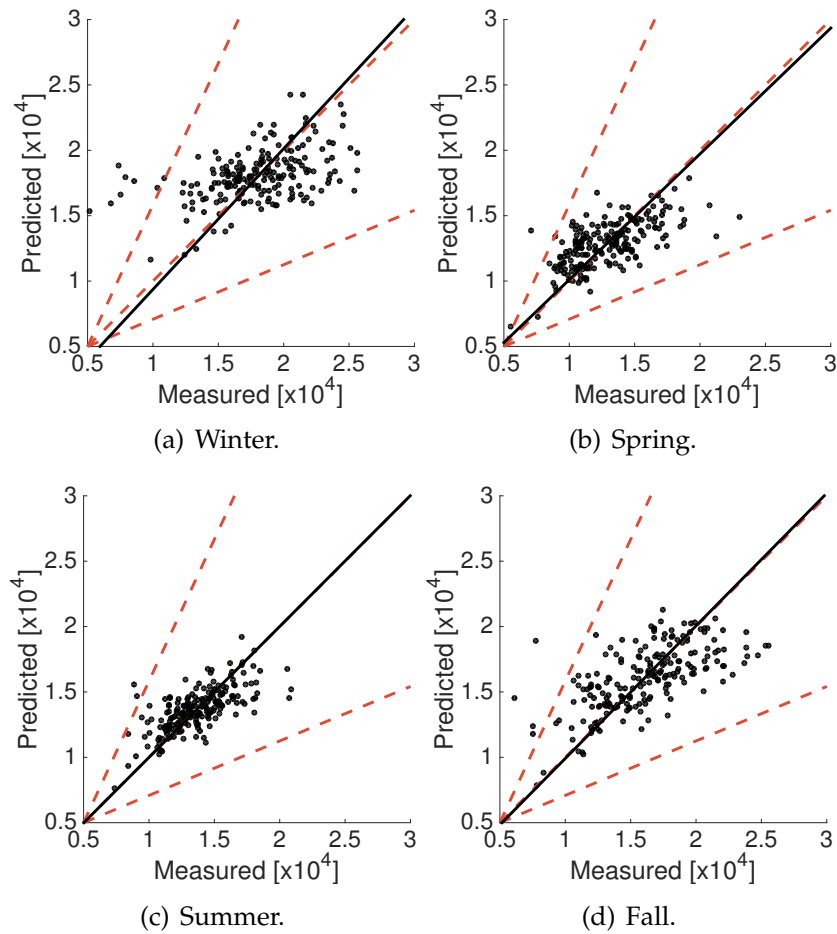


Figure 5.8: Scatter plots. Model predicted versus measured UFP concentrations ($\text{particles}/\text{cm}^3$) for seasonal pollution maps. The dashed lines depict the factor of two area (slopes of 0.5 and 2) and the 1-to-1 relation (slope of 1). All linear regression lines (solid) are very close to the 1-to-1 lines.

for particulate matter have R^2 values in 0.17–0.82 [HBdH⁺08, ZGT⁺14]. Most of these models are based on installations that are just in place for a short time (i.e., days or weeks) with measurement devices employed at a carefully selected set of around 40 locations, on average. In our deployment we have a large number of measurement locations, but (perhaps counterintuitively) with many locations it is more difficult to get predictions with high R^2 values [SBA⁺08, BRA⁺12]. We found that a restriction to 200 grid cells leads to a good trade-off among the considered metrics, as exemplified in Figure 5.10 for yearly pollution maps. Models created with a small number of grid cells have good R^2 s but their validity at locations not covered by the input cells is limited, which results in poor FAC2 and RMSE values. Models developed with many grid cells (> 300) have poor R^2 s and RMSEs. Among others, this is because the

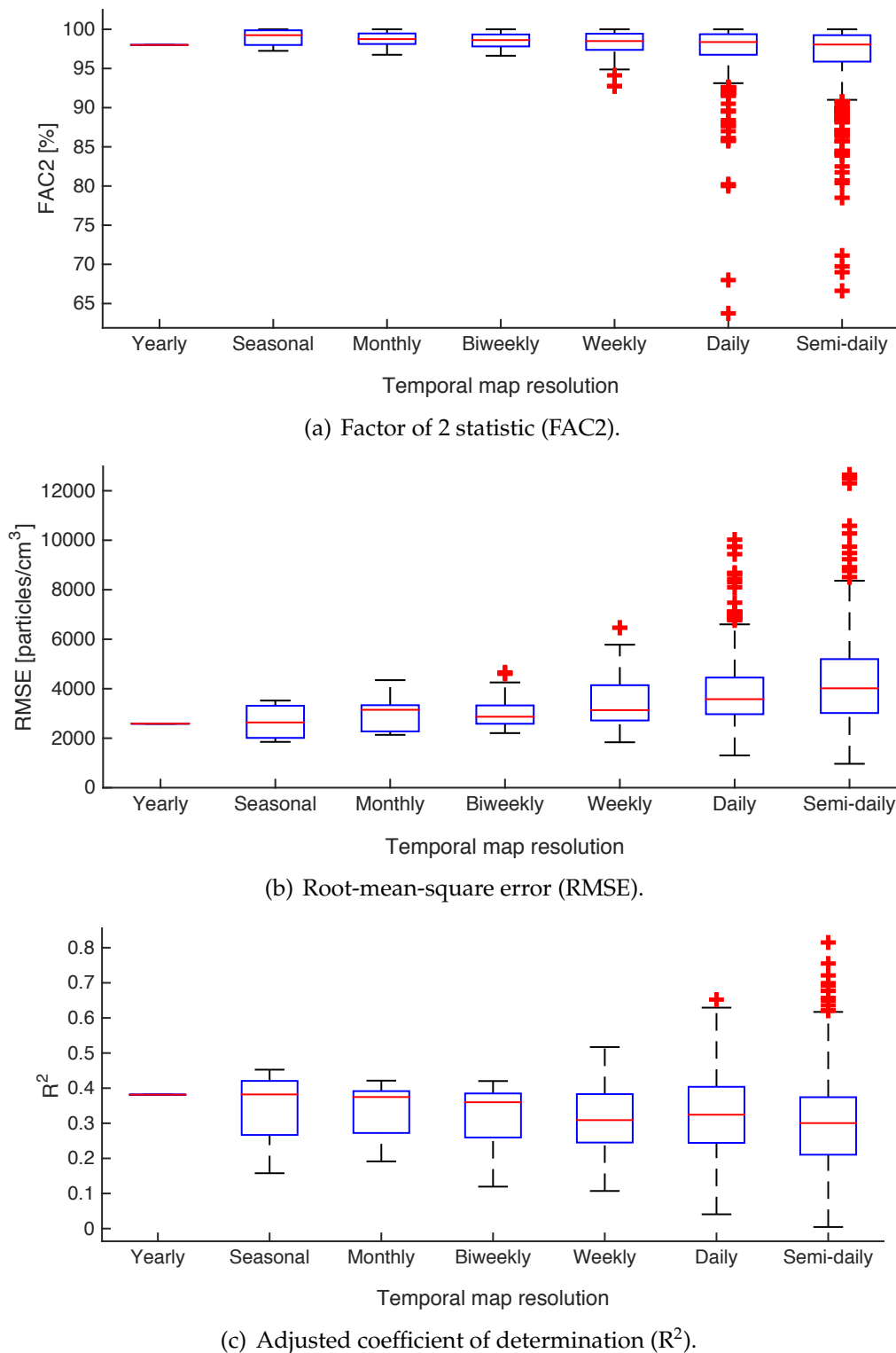


Figure 5.9: Models with yearly to weekly temporal resolutions have a good quality with high FAC2 values, low RMSEs, and acceptable R^2 coefficients. Models with higher temporal resolutions perform less well.

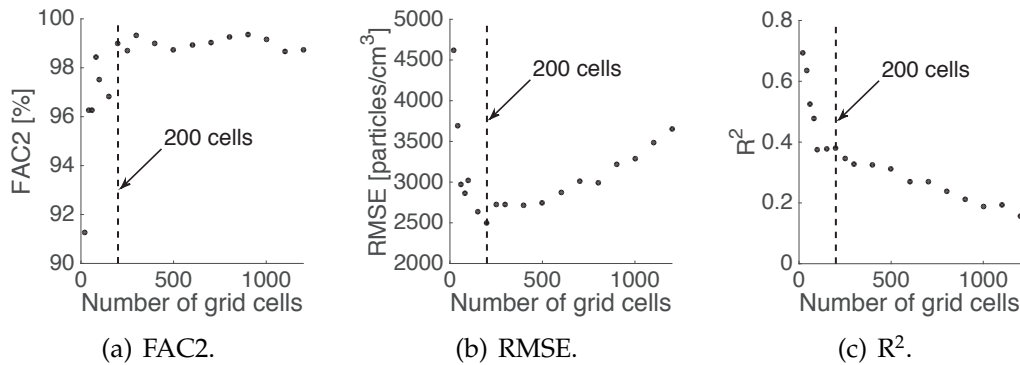


Figure 5.10: Yearly pollution maps modeled with 20–1200 grid cells. The models achieve a good performance trade-off with 200 cells (dashed line).

number of cells with unreliable average cell concentrations increases as the calculated means are more and more based on a limited number of measurements (remember that the cells are favored according to the number of measurements per cell).

5.5.3 Challenges in Developing Pollution Maps with High Temporal Resolution

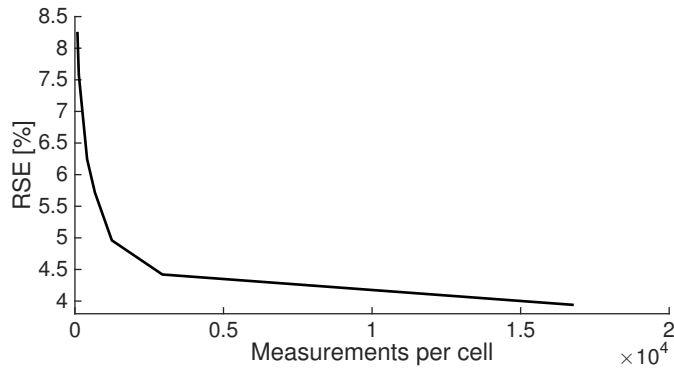
We observe on all performance metrics examined that for higher temporal resolutions it is more difficult to derive accurate pollution maps. The main problem is the reduced number of measurements available to calculate the mean particle concentration per cell, leading to the following complications:

- The calculated mean of a grid cell is less reliable as it is based on a smaller number of measurements.
- Erroneous and inaccurate measurements (e.g., due to outliers, sensor noise, and outdated calibration) have a higher impact on the mean value of a cell.

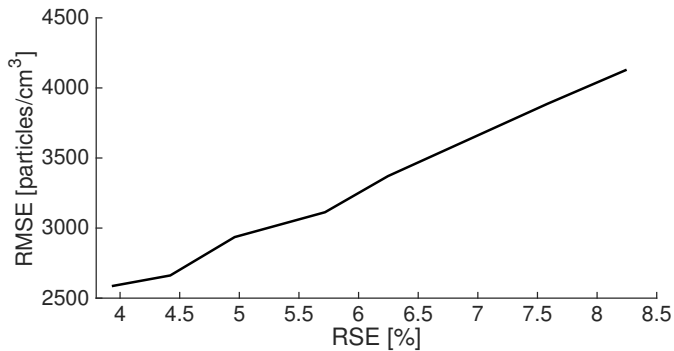
To quantitatively support our claims, we examine the relative standard error (RSE) of each cell, which is a good indicator of the reliability of a cell's mean:

$$\text{RSE}_i = \frac{\sigma_i^m}{\sqrt{n_i} \cdot c_i^m} \cdot 100, \quad (5.4)$$

where σ_i^m is the standard deviation of the n_i measurements of cell i . Figure 5.11(a) illustrates how the average RSE declines with increasing number of measurements per cell. Models derived from cells with small RSEs have lower RMSEs, as shown in Figure 5.11(b). The RSE increase from 4% to 8% results in a 60% rise of the models' RMSEs. Note that we



(a) RSE of the cells depending on the number of measurements per cell.



(b) RMSE of the pollution maps depending on the RSE of the cells.

Figure 5.11: With increased number of measurements the relative standard error (RSE) of the cells' mean decreases. Models based on cells with small RSEs have low RMSEs.

verified that using the median instead of the mean of a cell does not yield improved model performance.

In the following, we propose a new modeling approach to increase the number of measurements per cell and, therewith, improve the accuracy of models with high temporal resolution.

5.6 Increasing the Temporal Resolution Limit

We propose a novel modeling approach, which is able to make use of past pollution measurements to increase the accuracy of highly temporally resolved pollution maps. For that, we introduce a history database containing our measurements from the last 12 months annotated with metadata describing the environmental conditions at time of measurement. We exploit that concentrations of air pollutants show a high correlation with various environmental conditions (e.g.,

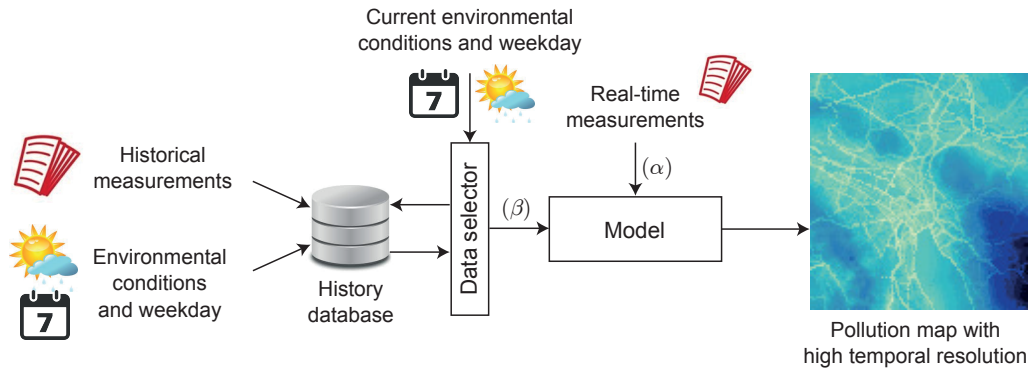


Figure 5.12: Additional data (β) from a database with historical measurements is used to enhance the original data set (α) to derive pollution maps with high temporal resolutions. The data selector ensures that only historical measurements are used, which were measured under similar environmental conditions as the average condition of the modeled time period.

temperature) and the current weekday [LPK09, CLCH⁺11, ZLH13]. The data selector depicted in Figure 5.12, selects from the history database those measurements that were performed on the same weekday and under similar environmental conditions as the average condition of the modeled time period. These historical measurements (β) are used together with measurements from the modeled time period (α) to calculate temporally resolved pollution maps, as shown in Figure 5.12. The enriched data set is used to calculate accurate pollution maps for short time scales. For example, assume we want to create a pollution map for a sunny but windy Saturday. We enrich the original measurements (α) from that Saturday with historical measurements (β) gathered on a past sunny, windy Saturday and use both data sets to calculate an accurate pollution map.

In the following, we show the feasibility of this novel approach on the case study of semi-daily pollution maps, where the standard model did not deliver a satisfactory accuracy.

5.6.1 Data Annotation

We feed the history database with 15 million UFP measurements from a complete year (April 2012 to March 2013). We annotate the measurements with the environmental conditions measured in the city center during this time period by a high-quality governmental station (the triangle in Figure 5.3 denotes the location of the station). The station provides 30 min averages of a diverse set of environmental variables, namely

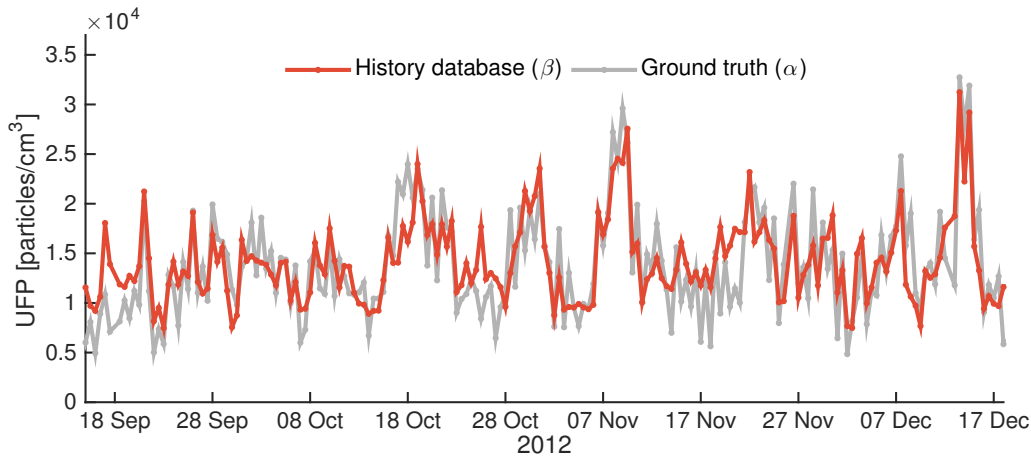


Figure 5.13: Average semi-daily concentrations based on the history database closely matches ground truth data measured on these days (Pearson $r = 0.74$).

humidity, atmospheric pressure, radiation, precipitation, temperature, wind direction, wind speed, and nitrogen oxide (NO_x) concentration.

5.6.2 Data Selector and Quality of Selected Data

The data selector, depicted in Figure 5.12, fetches based on the average environmental condition of the modeled half-day those measurements from the history database, which are most likely to be similar to the real measurements from that half-day. To find the best data selector, we evaluate all possible combinations of environmental variables. We introduce a deviation parameter, which controls how closely the metadata of the fetched measurements has to match the given conditions. For example, allowing 20% deviation and a temperature of 15°C , the data selector returns all data measured at $15^\circ\text{C} \pm 20\%$, whereas 100% refers to the maximum range in the database.

We evaluate the best data selector by comparing for a complete year the average semi-daily UFP concentration supplied by the two input data sets (β) and (α) depicted in Figure 5.12. We observe that the fetched data from the history database (β) are closest to the actual measured concentrations (α) if the selection is based on the following three criteria: temperature, wind direction, and NO_x concentration with an allowed deviation of 10%. The two data sets have a high Pearson correlation of 0.74 and a low average absolute difference of 2,500 particles/ cm^3 , as shown on a three-month extract in Figure 5.13. In this way, we attain on average a 14x increase in data volume from 19,000 (α) to 260,000 ($\alpha \cup \beta$) measurements per half-day.

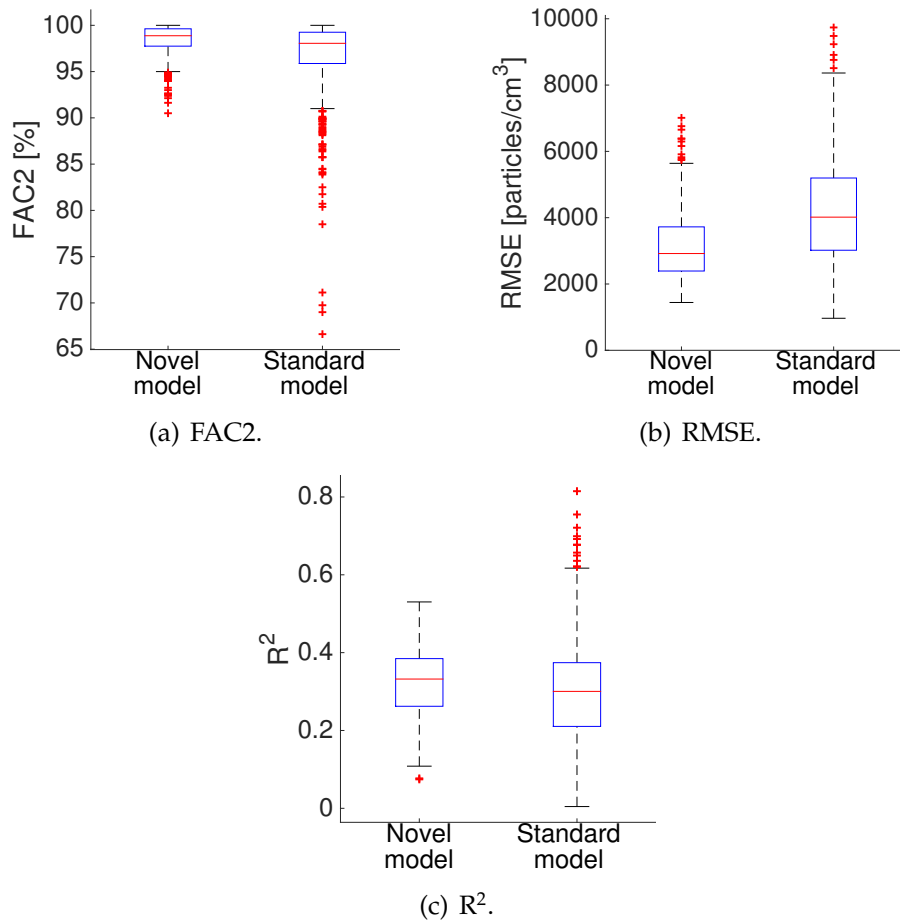


Figure 5.14: The novel method with a history database helps to increase the quality of semi-daily pollution maps. On average, the RMSE is 26 % lower using the novel model compared to the standard model.

5.6.3 Increased Quality of Semi-Daily Pollution Maps

The extended number of measurements helps to derive pollution maps with a higher accuracy. We compare our modeling approach with the standard approach using the three performance metrics. Figure 5.14 shows that using our novel modeling approach increases FAC2s by 2 %, R^2 s by 6 %, and decreases RMSEs by 26 %, on average. The semi-daily pollution maps created with the new models achieve a similar accuracy as the weekly pollution maps created with the standard models.

This new modeling approach advances the generation of accurate pollution maps with high temporal resolutions and simultaneously enables the forecasting of pollution maps for specific environmental conditions, e.g., by using weather forecast data to create pollution maps for the next hours or days.

5.7 Related Work

Monitoring airborne pollutants with mobile, low-cost sensors has gained much attention in recent years. Low-cost gas sensors are often embedded in custom-built, personal sensor nodes that are part of participatory sensing networks [HBPW08, CKCH09, DAK⁺09, CLL⁺14]. Participants can directly monitor the exposure level at locations where they spent their time [ZLW⁺14]. However, reaching the critical mass of users to get a coherent picture of the exposure situation in the area of interest is a formidable challenge and may require hundreds to thousands of contributors. Our approach uses a small number of non-personal sensor nodes deployed on top of public transport vehicles to automatically obtain a constant coverage in the area of interest, as described in Chapter 2.

There are different kinds of models that can be developed to predict intraurban pollution concentrations, such as proximity-based assessment, statistical interpolation, LUR, line dispersion models, and novel machine learning and data mining techniques [JAK⁺04, ZLH13, ZLW⁺14]. In this work, we develop LUR models since they have, compared to other models, a relatively low computational overhead, which is beneficial when deriving many hundreds of models. Above all, in the past, LUR models were applied to predict the concentration of a wide range of air pollutants [HBdH⁺08]. Most of these models are based on data from static monitoring stations installed at a small set of hand-picked locations (typically around 40) for some days or couple of weeks [HBdH⁺08]. In contrast, our mobile sensor nodes cover a much larger number of locations, i.e., every day around 300 grid cells of size 100 m × 100 m. Therewith, we develop maps with a high spatio-temporal resolution.

A number of studies address the challenge of developing models with high temporal resolutions. Liu et al. [LPK09] derive daily PM_{2.5} maps with a spatial resolution of 4 km by using satellite data in combination with meteorological features. Su et al. [SBA⁺08] extend their NO₂ and NO models with hourly meteorological conditions, such as wind speed and wind direction, to create pollution maps with an hourly resolution. However, it remains unclear how much the meteorological data helps to improve their models. Furthermore, the performance of the hourly resolved pollution maps was not evaluated.

Clifford et al. [CLCH⁺11] and Mølgaard et al. [MHCH12] develop UFP models with high temporal resolutions by using not only traffic and land-use characteristics data but also meteorological variables as model covariates. Both works solely model temporal variability as their data originates from a single static station.

Many of the above works use meteorological data to enhance their models, i.e., [LPK09, SBA⁺08, CLCH⁺11, MHCH12]. However, their

methods of integrating meteorological data into the modeling process only work if the *(i)* specific meteorological conditions are known at each measurement location and *(ii)* meteorological conditions are significantly different among locations for a given instant in time. Most often, both conditions do not apply for spatially resolved intraurban data sets.

5.8 Conclusions

Today, little is known about the spatial distribution of air pollutants in urban environments. We address this problem by using one of the largest spatially resolved ultrafine particle data set available today collected with our mobile air pollution monitoring network. We develop land-use regression models to create pollution maps with a high spatial resolution and study their temporal resolution limit. We reveal that the accuracy of pollution maps with sub-weekly temporal resolution suffers from the limited number of measurements available to model the pollution concentrations. We tackle this problem by proposing a novel modeling approach, which is able to make use of past measurements to increase the available data volume. Therewith, we develop accurate ultrafine particle pollution maps with a high spatio-temporal resolution. These maps are valuable to the general public as well as to environmental scientists and epidemiologists to shed more light on the adverse health effects of ultrafine particles. We believe that our findings promote the accurate, detailed, and timely assessment of air quality in urban environments.

In Chapter 6, we use the derived high-resolution pollution maps to enable a novel route planner application, which helps urban dwellers to assess and reduce their exposure to ultrafine particles.

6

The Health-Optimal Route Planner

High-resolution pollution maps enable a wide range of new application scenarios, which are not feasible with pollution maps available today. We end the thesis by demonstrating an interesting use case. We combine the pollution maps computed in Chapter 5 with the street network of the city of Zurich to analyze by how much urban dweller can reduce their exposure to air pollution by not taking the shortest path between origin and destination but a healthier and slightly longer alternative route. We introduce a new weight function to assess the exposure on each street segment and evaluate the benefits of the healthier path on 5000 randomly generated source-destination pairs. Finally, we efficiently implement the algorithm in an easy to use smartphone application to help city residents to understand and reduce their exposure to air pollutants.

6.1 Introduction

Many companies offer route planning services (e.g., GoogleMaps [Goo14], MapQuest [Map14], and TomTom [Tom14]), which evaluate the best route between origin and destination based on different criteria, such as distance, journey time, and least total fuel consumption, for various transportation modes. However, the general public, especially pedestrian and cyclist commuters, are increasingly concerned about the adverse health effects of urban air pollution and, hence, wish to minimize their exposure to airborne pollutants [DHBNH10, CLL⁺14]. Up to today, the lack of spatial resolved pollution data hinders route planners to provide such air pollution related services.

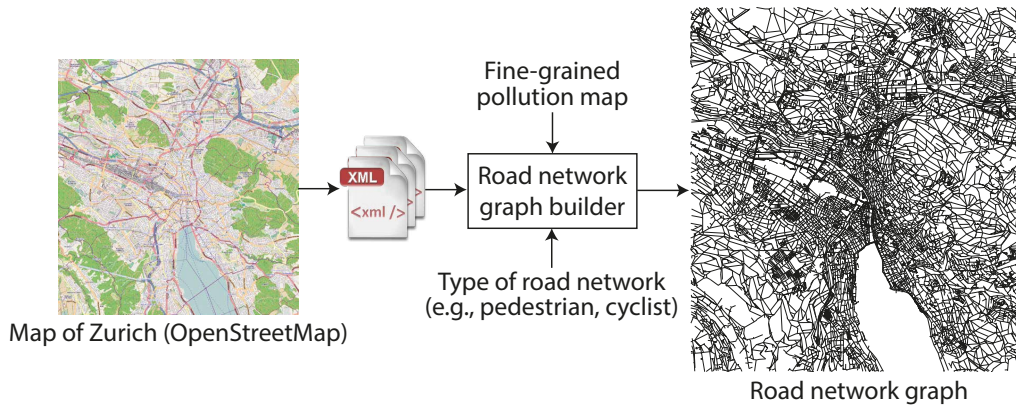


Figure 6.1: The road network builder uses OpenStreetMap data to create different types of road network graphs, where nodes represent street crossroads and dead-end streets and edges connect two consecutive nodes.

We use the generated high-resolution air quality models from Chapter 5 to derive in Section 6.2 a novel cost function for all street segments of Zurich’s road network. The cost function expresses the expected exposure to UFPs when traversing the respective street segment. We implement this new cost function in a pathfinding algorithm, described in Section 6.3, to compute health-optimal routes and compare them in Section 6.4 to the traditionally used shortest path routes. Therewith, we analyze the potential of reducing the number of inhaled particles while roaming the city. Finally, we implement in Section 6.5 the health-optimal route planner as iPhone app, which is efficiently running on even the oldest smartphone models supported. We survey related work in Section 6.6, and conclude in Section 6.7.

6.2 Road Network Graph

We use data from OpenStreetMap to construct a graph of Zurich’s road network.¹ We export the area of Zurich in the OSM XML file format and process the data with our road network graph builder to create a graph of Zurich’s road network, as depicted in Figure 6.1. Depending on which type of road network is requested (e.g., for pedestrians, cyclists, or car drivers), a different graph is generated.

We represent the road network by an undirected graph $G = (V, E, W_d, W_p)$ comprising a set V of vertices (nodes), a set E of edges each connecting a pair of nodes, and two sets of edge weights W_d and W_p . Each node $v_i \in V$ represents a crossroad or a dead-end street of

¹<http://www.openstreetmap.org>

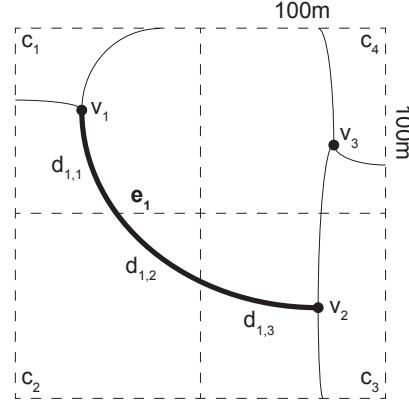


Figure 6.2: Road segment e_1 connecting the two nodes v_1 and v_2 penetrates three 100 m x 100 m grid cells with a corresponding pollution exposure of c_1 , c_2 , and c_3 .

the road network. There is an edge $e_j \in E$ between two nodes of V if the nodes are directly connected by a road segment (i.e., without a crossroad in between). Each edge is associated with two weights. Weight $w_{d,j} \in W_d$ denotes the length of the road segment between the pair of nodes connected by edge e_j . Weight $w_{p,j} \in W_p$ denotes the expected pollution exposure on the road segment represented by edge e_j . We assume that the travel time between two nodes of an edge $e_j \in E$ is proportional to the distance $w_{d,j}$. The number of particles a human inhales is proportional to the time of exposure [KHB⁺03]. Hence, we multiply the modeled particle concentration along edge e_j with the length of the corresponding road segment:

$$w_{p,j} = \sum_{k=1}^n d_{j,k} \cdot c_k, \quad (6.1)$$

where n is the number of 100 m x 100 m grid cells the edge penetrates, $d_{j,k}$ is the length of edge e_j penetrating grid cell k , and c_k is the modeled pollution concentration in grid cell k . Note that the sum of the n subsegments is equal to the total length of the road segment, i.e., $w_{d,j} = \sum_{k=1}^n d_{j,k}$.

We illustrate the above weight functions on a simple example depicted in Figure 6.2. The two crossroads v_1 and v_2 are connected by road segment e_1 , which cuts through three grid cells with modeled pollution concentrations c_1 , c_2 , and c_3 . We calculate the weights $w_{d,1}$ and $w_{p,1}$ as:

$$w_{d,1} = d_{1,1} + d_{1,2} + d_{1,3}, \quad (6.2)$$

$$w_{p,1} = d_{1,1} \cdot c_1 + d_{1,2} \cdot c_2 + d_{1,3} \cdot c_3. \quad (6.3)$$

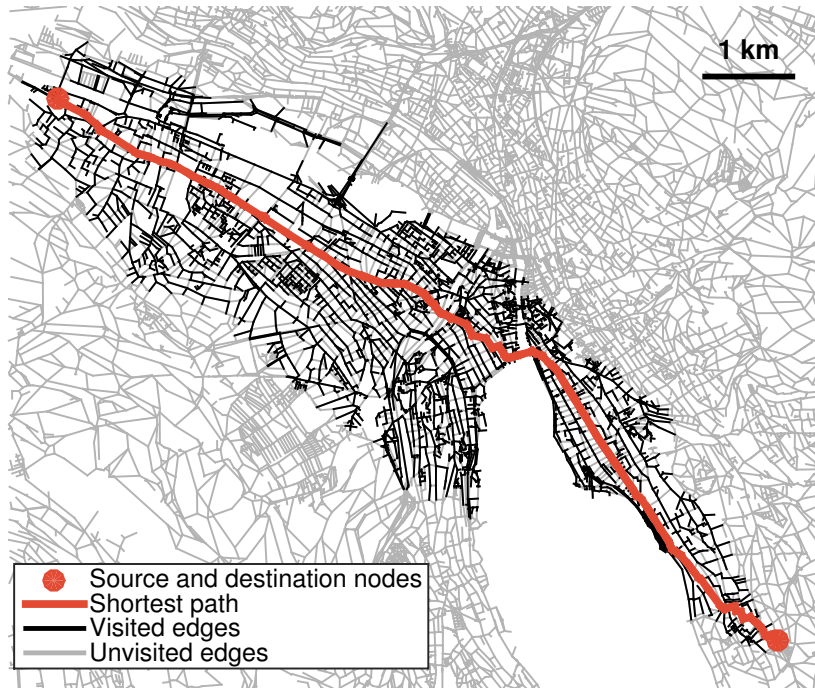
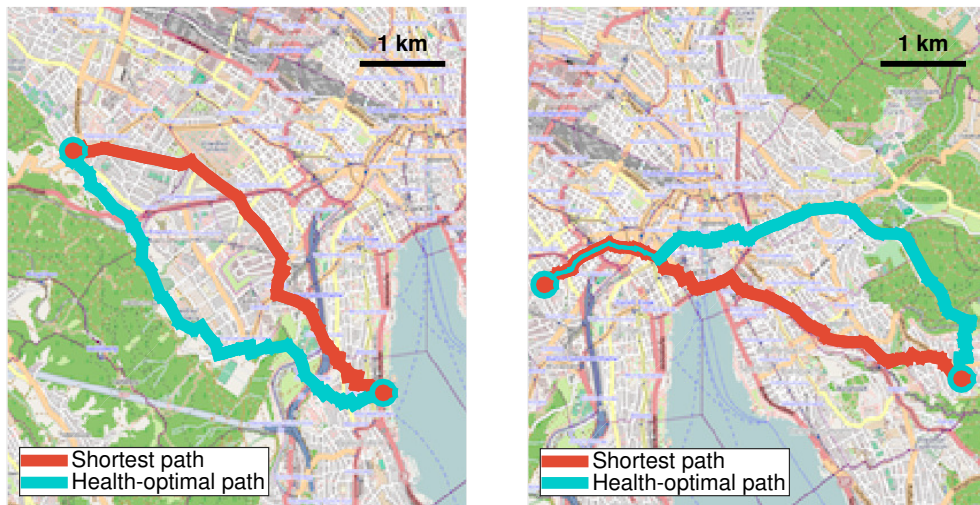


Figure 6.3: The A* pathfinding algorithm makes use of a heuristic estimation of the cost involved to reach the destination node, enabling a goal-directed exploration of the road network graph.

6.3 Finding the Least-Cost Path

We use all road segments available (including those only accessible by pedestrians) to create graph G representing the road network of Zurich consisting of $|V| = 27,000$ nodes (i.e., crossroads and dead-end streets) and $|E| = 74,000$ edges. We implement the widely used A* pathfinding algorithm [HNR68] to find the least-cost path between two nodes of the road network with respect to the two weight metrics introduced. A* requires for every node $v_i \in V$ a heuristic estimate $h(v_i)$ of the cost involved to reach from node v_i the destination node. A* only guarantees optimality of the computed path, if an admissible heuristic is used, i.e., if the heuristic never overestimates the cost of reaching the destination node. To ensure optimality, we use line-of-sight distance as heuristic when computing the shortest path. To compute the health-optimal path (with respect to UFP exposure), we use the product of line-of-sight distance and minimal pollution concentration within the modeled area. With the help of these admissible heuristics, A* is able to perform a goal-directed exploration and quickly find the least-cost path, as illustrated in Figure 6.3 for the shortest path between a pair of source-destination nodes.



(a) Albisrieden to Wollishofen.

(b) Wiedikon to Witikon.



(c) Höggerberg to ETH Zentrum.

Figure 6.4: Three exemplary source-destination pairs within the city of Zurich, Switzerland. In (a) the shortest and health-optimal routes are completely different while they partly overlap in (b) and (c).

6.4 Comparing Shortest Paths to Health-Optimal Paths

In the following, we study how much urban dwellers can reduce their pollution exposure by not taking the shortest but the health-optimal path between two arbitrary locations in the city. We randomly pick from the road network graph 1000 source-destination pairs. We require a minimum straight-line distance of 5 km between source and destination nodes to prevent very short paths where with high probability the shortest and

health-optimal paths are identical. For these 1000 source-destination pairs we compute for both introduced edge weights the least-cost paths using the A* pathfinding algorithm. Then, we compare the length and the pollution exposure of the two routes. To evaluate the pollution exposure, we use the seasonal UFP model depicted in Figure 5.7(d), which is based on our pollution measurements from October to December 2012.

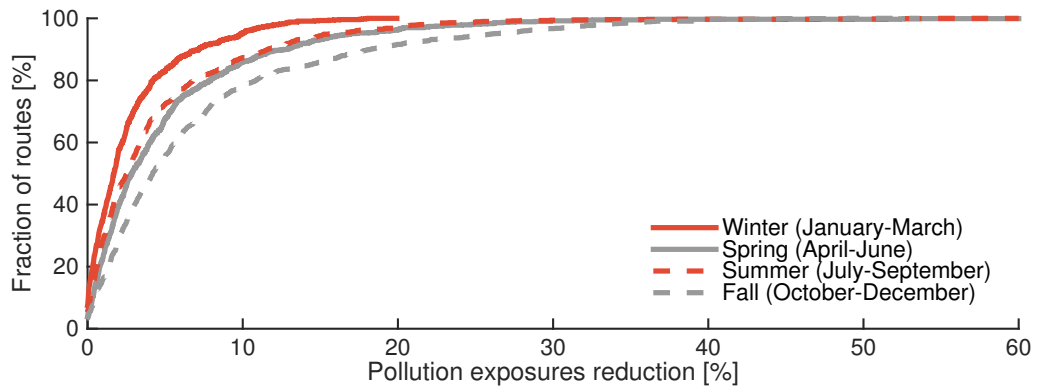
We find that taking the health-optimal path instead of the shortest path yields an exposure reduction of 7.1 % ($6.8 \cdot 10^6$ particles/cm³·m) on average. This comes at a price of longer travel distances, the length of the health-optimal path is 6.4 % (548 m) longer on average. However, the difference between the shortest and health-optimal paths largely depend on the given source and destination nodes, as illustrated in Figure 6.4. The computed routes are completely different in Figure 6.4(a), partly overlap in Figure 6.4(b), and are almost identical in Figure 6.4(c). The additional distance (and hence exposure time) incurred by not following the shortest path only pays off if significantly less polluted road segments are available nearby. We can clearly observe this in the depicted health-optimal paths in Figure 6.4, which if possible pass through or are close to public green spaces characterizing areas with low pollution concentrations.

The achievable reduction also highly depends on the specific distribution of the air pollutants. We reuse the 1000 source-destination pairs from above to compare the shortest paths with the health-optimal paths for all four seasonal air quality models depicted in Figure 5.7. We find that during the fall months the reduction in pollution exposure is more significant than during the winter months, as depicted on the CDF plots in Figure 6.5. This is because during the fall months the pollution concentration along major roads is considerably higher than on smaller streets close by, as observable in Figure 5.7(d). In contrast, during winter months this pollution gradient is smaller.

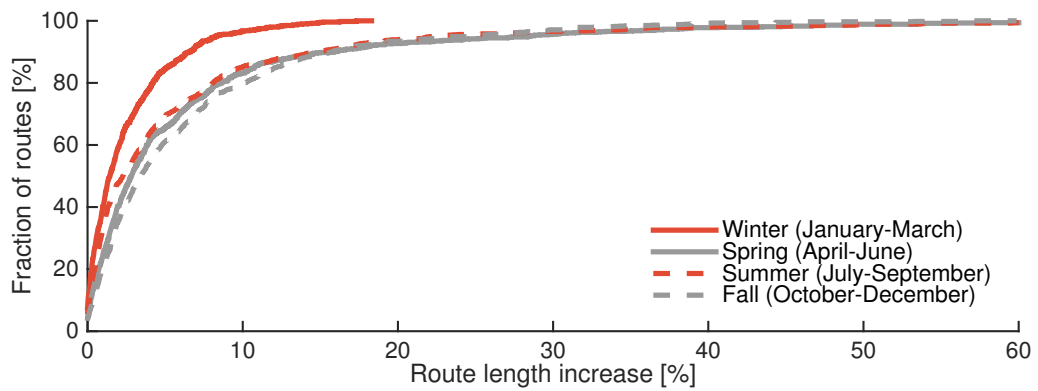
6.5 Smartphone Application

We implement the application in iOS for iPhones and iPads, as depicted in Figure 6.6, and make it available for free on the App Store as *hRouting–The Health-Optimal Route Planner*.² On the start screen, depicted in Figure 6.6(a), the user can enter its choice of origin and destination. The addresses are automatically autocompleted and it is checked whether they reside in the region supported (wider area of Zurich). Upon pressing "Compute route", the application computes in parallel (all recent

²hRouting–The Health-Optimal Route Planner: <https://itunes.apple.com/us/app/hrouting-health-optimal-route/id931299863>



(a) Reduction in pollution exposure.



(b) Increase in route length.

Figure 6.5: The health-optimal paths minimize the expected exposure to UFPs while being only slightly longer than the shortest paths. The differences between the health-optimal and shortest paths are most significant in the fall months.

iPhones have a dual-core processor) the shortest (red) and the health-optimal (green) routes and displays them on top of Google Maps, as shown in Figure 6.6(b). Clicking the info button in the top right corner displays additional information about the shortest and health-optimal routes and their differences in length and air pollution exposure, as illustrated in Figure 6.6(c). The History lists routes computed in the past, the Settings tab displays user preferences, and the About tab offers background and contact information (see Figures 6.6(d)–(f)).

To be accepted by the users, it is essential that the smartphone application is reactive. Long computation times must be avoided as much as possible. Therefore, it is important to efficiently implement the least-cost pathfinding algorithm. We achieve this by using simple C data structures (mainly integer arrays and priority queues) and primitive C data types instead of the more sophisticated, but higher overhead inducing, object-oriented data structures of Objective-C. We time the

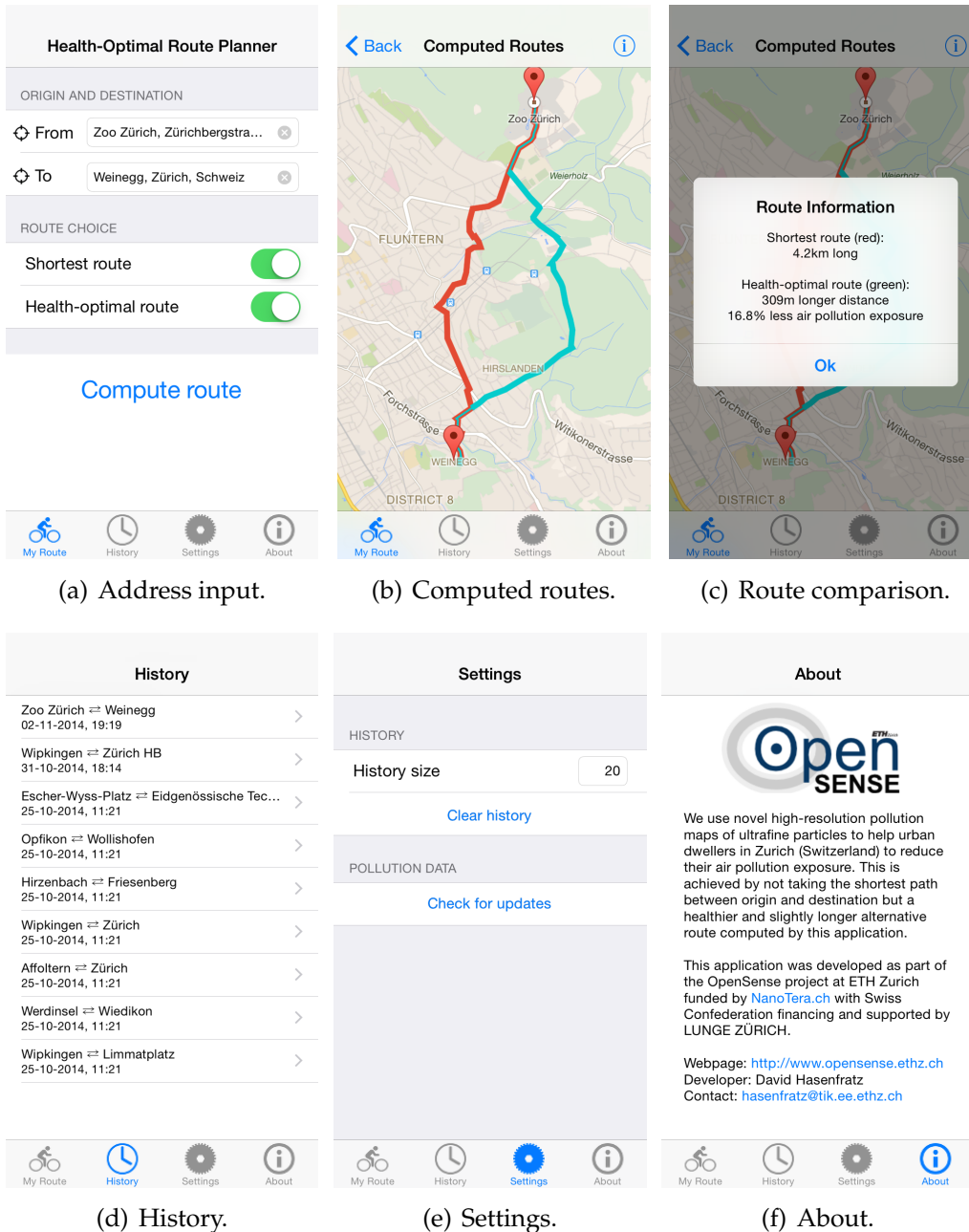


Figure 6.6: The iPhone and iPad application computes for given origin and destination addresses the shortest (red) and health-optimal (green) routes. The routes are displayed on top of Google Maps and the user gets detailed information about the length and pollution exposure differences.

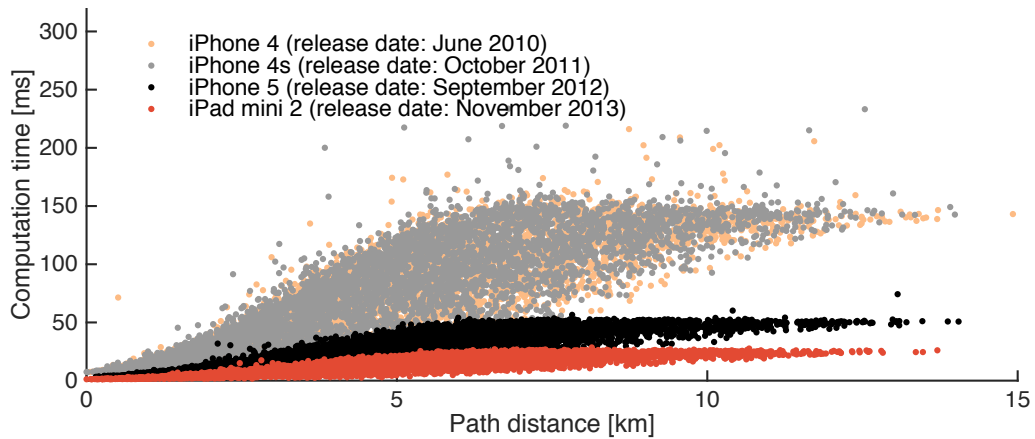


Figure 6.7: Least-cost path computation times for 5000 source-destination pairs on an iPhone 4 and iPhone 4s (both on average 80 ms), iPhone 5 (on average 28 ms), and iPad mini 2 (on average 14 ms).

execution times for 5000 random source-destination pairs. We minimize the interference of other applications and the operating system by ensuring that no other application is running and by putting the phone in airplane mode (suspends all signal transmitting functions of the device). We achieve to compute a route within 28 ms on an iPhone 5 (release date September 2012) on average, as depicted in Figure 6.7. The older iPhone 4 and 4s (release date June 2010 and October 2011) have considerably longer computation times with 80 ms on average, while the newer tablet iPad mini 2 (release date November 2013) is very fast with an average computation time of 14 ms. This shows that even the oldest iPhone generation supported by our application is able to compute the majority of routes in the benchmark set in less than 150 ms.

6.6 Related Work

High-resolution pollution maps enable a wide range of novel applications. Cheng et al. [CLL⁺14] propose two web and six smartphone applications, including a service to get the pollution exposure at the user’s current location and a virtual AirPet, which lives in the user’s phone and grows up healthy and happy if fed at locations with good air and unhealthy and sad otherwise.

Several recent studies have analyzed how the choice of route impacts exposure to air pollutants. Hertel et al. [HHK⁺08] analyze the commute of 50 persons and find that a careful route selection can significantly reduce the air pollution exposure. Due to the lack of spatially resolved

air pollution data, diurnal traffic is used as a proxy for air pollution exposure. Sharker et al. [SK14] use data from sparsely distributed static measurement stations to estimate the pollution level of each street segment of the analyzed road network. The weight of a street segment is calculated as product of the estimated pollution level and expected travel time through the street segment. Cheng et al. [CLL⁺14] propose a web-based application to compute paths, which minimizes the exposure to PM_{2.5}. The authors collect data with ten fixed and two mobile air quality monitoring devices for two months in an 4 km x 4 km area and use an interference model based on Gaussian process regression to compute an exposure heat map. Finally, Allemann et al. [AR13] outline the health-atlas platform, a route planing service that plans to incorporate different modes of transportation, personal mobility restrictions, and air pollution measurements (governmental and crowd-sourced). This should facilitate the system to suggest its users low exposure routes between origin and destination.

6.7 Conclusions

High-resolution maps enable a wealth of new application opportunities. We use the ultrafine particle maps created with the measurements from our mobile air pollution monitoring network to build the iOS application *hRouting—The Health-Optimal Route Planner*. Users can enter origin and destination addresses and let the application compute and display the shortest and a slightly longer but healthier alternative route. We analyze on a random set of source-destination pairs by how much urban dwellers can reduce their exposure to air pollutants. We find that the air pollution exposure can be reduced by up to 60%. On average the exposure is 7% lower if not the shortest but the health-optimal route is taken, which is on average 6% longer than the shortest route.

7

Conclusions and Outlook

This chapter summarized the contributions of this thesis and outlines possible future research directions.

7.1 Contributions

Air pollution has become the world's single biggest environmental health risk. Atmospheric pollutants considerably affect human health; they are responsible for a variety of respiratory and cardiovascular illnesses and are known to cause cancer. Additionally, air pollution is responsible for environmental problems, such as eutrophication of ecosystems.

Nowadays, air pollution is monitored with a small number of highly accurate but fixed measurement stations. As a result, very little is known about the spatial distribution of air pollutants. However, the availability of reliable intraurban pollution maps is essential. They raise the citizens' awareness for air pollution, empower environmental scientists to craft and evaluate new policies, and help epidemiologists to better understand the adverse health effects of air pollution.

In this thesis, we tackle this challenge and derive accurate intraurban pollution maps with an unprecedented spatio-temporal resolution. The basis forms our air pollution monitoring network consisting of mobile, low-cost sensor nodes. We introduce algorithms, which use knowledge gathered at meeting points between sensors (mobile and static), to increase measurement quality of the deployed sensors, which have due to size and cost constraints a limited stability, sensitivity, and precision. We use the processed measurements to derive pollution maps with a high

spatio-temporal resolution. These maps are useful to get insights about the spatial distribution of air pollutants opening up a wealth of new application scenarios. We believe that our findings serve as a stepping stone towards detailed real-time pollution assessment.

The main contributions of this thesis are:

- **Mobile air pollution monitoring network (Chapter 2)**

Based on existing hardware and software infrastructures, we build a mobile air pollution monitoring network using low-cost sensors enabling us to monitor various airborne pollutants and environmental parameters with a high spatial resolution. The sensor nodes are installed on top of ten streetcars of the public transport network. To the best of our knowledge, over the past three years these sensor nodes collected the largest spatially resolved urban air quality data set available today.

- **Multi-hop calibration of mobile sensor networks (Chapter 3)**

Many low-cost sensors need periodic re-calibrations to deliver measurements with a good data quality during the whole system lifetime. We are the first to study multi-hop calibration (with respect to a reference signal) in detail. We introduce a multi-hop calibration algorithm using geometric mean regression, which highly reduces error propagation in the network. Our algorithm is especially valuable when calibrating large networks of heterogeneous sensors.

- **Data quality assessment of low-cost sensor readings (Chapter 4)**

Low-cost sensors often have a reduced measurement accuracy and reliability. We introduce a model-driven approach to provide reliable accuracy bounds for individual noisy sensor readings. Further, we detect measurement outliers and sensor failures. We achieve this by integrating generic models to describe the phenomenon of interest and the sensor characteristic.

- **Intraurban pollution maps and its applicability (Chapters 5 and 6)**

We use the measurements collected by our air quality monitoring network to derive intraurban pollution maps. We increase the quality of high-resolution pollution maps by using a new modeling approach, which incorporates past measurements annotated with metadata, such as environmental conditions, into the modeling process. In this way, we derive accurate pollution maps with a high spatial and temporal resolution. These maps facilitate new applications, e.g., a novel health-conscious route planning service.

7.2 Possible Future Directions

The work presented in this thesis represents an important step towards reliable high-resolution assessment of urban air pollutants. Nevertheless, there exists potential for further improvements and extensions. The following list details possible directions for future research:

- **Automatic sensor failure detection**

Sensor failure detection is crucial, especially in a multi-hop calibration scenario, where wrong sensor readings can jeopardize the calibration of many other sensors in the network. This is particularly a problem, if the faulty sensor is close to the root of the network, i.e., close to the reference sensor initiating the calibration process. We are able to detect some erroneous measurements with the proposed correlation threshold, which ignores meeting points between sensors if the correlation of their measurements is below a given threshold. A promising idea for faulty sensor detection would be to look at the overall calibration quality of the sensor network depending on the subset of sensors calibrated. If a faulty sensor is removed from the network, the calibration quality of the remaining sensors should increase. In contrast, if a correct sensor is removed then the calibration quality of the network should decrease, as, in general, fewer meeting points between sensors result in a poorer calibration of the sensors.

- **Calibration of static sensor networks**

The calibration algorithm proposed in this thesis exploits meeting points between mobile and static sensors to enhance the calibration of low-cost sensors. Within a given time slot, only those sensors are calibrated, which are part of the rendezvous graph, i.e., sensors, which were in the spatial and temporal vicinity of other static or mobile sensors in the given time slot. Hence, our method can not be used to calibrate static sensor networks, e.g., the community-led Air Quality Egg sensing network.¹ Therefore, a very interesting and valuable extension would be to enable the calibration of static sensor networks. One possibility is to identify points in time where sensors are exposed to similar conditions despite being at different locations, e.g., during the night the concentration of many pollutants drops towards zero. Such kind of opportunities could be used to also calibrate sensors, which are never in the vicinity of other sensors.

¹<http://airqualityegg.com>

- **On-the-fly sensor calibration**

In this thesis, we use a slotted approach to calibrate sensors in the network. It would be interesting to calibrate sensors on-the-fly enabling the gathering of accurate measurements in real-time. However, on-the-fly calibration introduces additional challenges. For example, when two sensors meet a metric is needed to decide which sensor is more accurately calibrated and, hence, can serve as calibration parent for the other sensor. How to define such a metric is a challenging open question.

- **Fine-grained pollution maps and its application areas**

There are different kind of modeling approaches, which can be used to predict intraurban pollution concentrations, e.g., statistical interpolation, land-use regression, and line dispersion models. In this thesis, we apply land-use regression models, which are widely used to assess the spatial variability of air pollutants. However, it would be interesting to use our unique data set to compare different modeling approaches to analyze their pros and cons in modeling the distribution of air pollutants in urban areas.

High-resolution pollution maps enable in combination with other spatial data sets (e.g., population density, health-insurance premiums, and real estate prices) many new interesting possibilities, which are to be explored.

- **Participatory sensing: trust, privacy, and energy efficiency**

The methods proposed in this thesis for calibration, assessment of data quality, and air quality modeling are all well suited to be used on large data sets gathered by participatory sensing networks. However, incorporating participatory sensed data poses additional challenges for the introduced algorithms, not covered in this thesis:

- *Trust*

A mechanism is needed to decide whether the sensor readings of a user are trustworthy. If the sensor readings are deemed as wrong, we need a mechanism to classify the cause to react accordingly. For example, users delivering intentionally wrong measurements should be banned, measurements performed under improper circumstances (e.g., while device is in the pocket) should be ignored, and users gathering data with erroneous or uncalibrated sensors should be informed such that they can take counteractions. Distinguishing these different scenarios is a challenging task.

– *Privacy*

In general, users need to share location information. However, often the absolute measurement location is not required and location obfuscation mechanism could be used to retain user privacy. For example, land-use regression models do not need the coordinates of the measurement location but only require the land-use and traffic characteristic of the respective location.

– *Energy-efficiency*

In many participatory sensing scenarios energy-efficiency is key. The measurement device should automatically decide, based for example on the device's energy state and calibration and air quality model requirements, when it is best to perform measurements, e.g., locations already covered by other users are less interesting than locations not visited by anybody.

A

Monitoring Radio-Frequency Electromagnetic Fields

Radio-frequency electromagnetic fields are emitted by many applications, such as radio broadcasting and mobile communication. A part of the general public is increasingly concerned about the long-term effects of electromagnetic radiation on the human health. However, the accurate exposure assessment in people's everyday life remains a formidable challenge. State-of-the-art personal exposure meters are expensive and tedious to use. Epidemiological large-scale studies are rare and governmental compliance measurements can only cover a small number of locations of high interest (e.g., schools and hospitals).

In this appendix, we demonstrate that accurate, spatially resolved electromagnetic field measurements are feasible with commodity sensor nodes of our streetcar deployment introduced in Chapter 2. We collect over 4 million radio-frequency electromagnetic field measurements and use them, by applying the methods described in Chapter 5, to develop the first GSM exposure map of Zurich with a spatial resolution of 100 m. Further, we compare the found exposure levels to measurements from different urban cities across Europe.

A.1 Introduction

Wireless communication based on radio-frequency electromagnetic fields (RF-EMFs) is omnipresent in our everyday lives. Besides the well-established radio and television broadcastings, there is a massive increase

of new devices using wireless technologies such as cell phones, cordless telephones, WiFi capable computers, and short-range communication technologies. With the advent of these relatively new technologies during the last decade, part of the general public raised concerns about the long-term effects of RF-EMFs on human health, e.g., [ETB⁺10].

In general, the measured strengths of RF-EMFs in urban environments are considerably below the exposure limits that are known to be hazardous [ABB⁺98]. However, many effects of RF-EMF radiation on the human body are not well understood because there is a lack of long-term and large-scale epidemiological studies quantifying these effects [FMBF⁺12].

A broad field of medical studies examine the influence of RF-EMF radiation on the human health. Examples of such investigations are evaluations on the influence of RF-EMFs on early childhood cancer [ETB⁺10] and brain tumor [KTK⁺09]. Many of these studies have in common that they rely on weak proxies (e.g., distance to closest cell tower [ETB⁺10]). However, the relation between exposure level and distance to the source is complex, especially in urban areas.

Exposure meters developed for personal measurements are used in medical studies [FMB⁺10] and for governmental supervision [Off14] to measure the strengths of RF-EMFs. The handling of these exposure meters, such as the EME Spy 120 (Satimo) [EME14] and ESM-140 (Maschek) [Mas14], is an elaborate and time-consuming task. As a result, nowadays periodic measurements are only available from specific spots of high interest, e.g., schools, playgrounds, and hospitals. For example, in the city of Zurich seven spots were observed by the government from 2009 to 2012 with on average two measurement campaigns per spot, each lasting for 30 min [Off14].

Contributions and road-map. Many of today's gadgets are equipped with a wide range of different wireless technologies, e.g., WiFi, FM radio broadcasting, and Bluetooth. Instead of just using the wireless modules of these devices for communication, we propose to use them for RF-EMF exposure monitoring on the supported frequency bands. Therewith, it is not possible to monitor as many frequency bands as with traditional exposure meters. However, the frequency bands that can be covered with commodity hardware, such as GSM and UMTS (downlink), WiFi, and FM, are the most interesting ones as they are *large contributors* to the total field strength exposure in urban areas [FMN⁺09]. Furthermore, these measurements are valuable because they reveal the background exposure levels, which are omnipresent and can not be influenced by personal behavior (such as less frequent phone usage). So far no simple proxies exist to precisely model these background exposure levels [BFT⁺10].

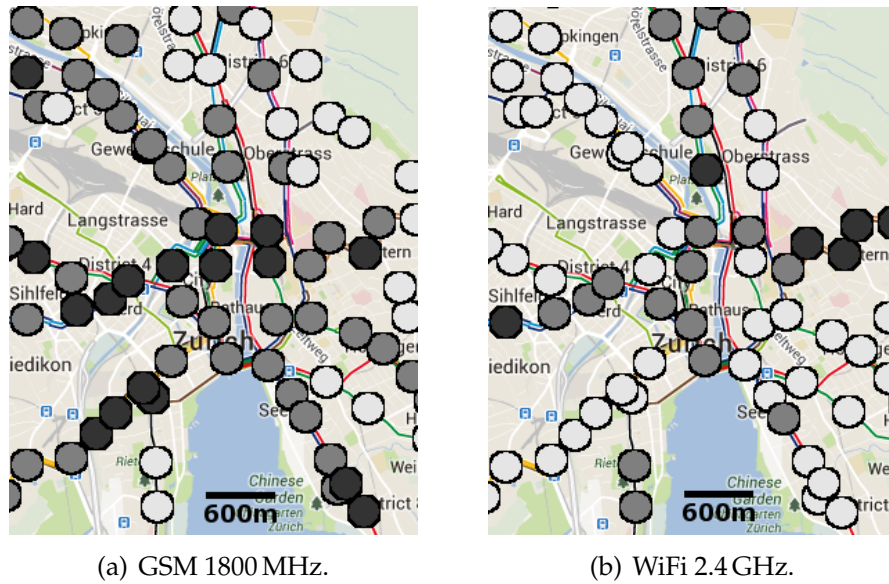


Figure A.1: Electromagnetic exposure levels in the city center of Zurich (Switzerland) from GSM and WiFi frequency bands. High (black), medium (gray), and low (light gray) exposure levels are distinguished.

We present the design, implementation, and evaluation of our mobile measurement system, which facilitates long-term, fine-grained, and fully automatic RF-EMF monitoring. We use our mobile streetcar sensor network introduced in Chapter 2 to collect RF-EMF measurements throughout the city of Zurich, as depicted in Figure A.1. The main contributions of this appendix are:

- We demonstrate that accurate RF-EMF exposure measurements are feasible with commodity sensor nodes. Our measurement accuracy is similar to state-of-the-art exposure meters widely used in medical studies and to control the adherence to exposure limits.
- We present first insights from our monitoring campaign. For instance, we reveal a 9-fold exposure increase over the past four years on the 2.4GHz WiFi band at urban outdoor locations and develop the first high-resolution GSM exposure map of Zurich.

We provide in Section A.2 a brief introduction to electromagnetic fields and describe our system design in Section A.3. In Section A.4 we show the accuracy of our measurements by comparing them with reference data from a state-of-the-art personal exposure meter and present first results from our streetcar deployment in Section A.5. We survey related work in Section A.6, and conclude in Section A.7.

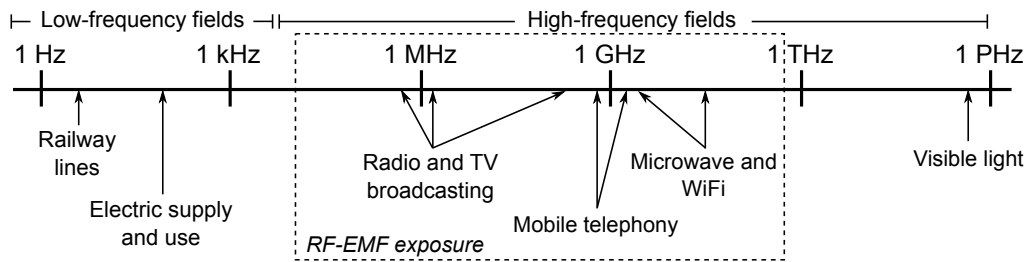


Figure A.2: Electromagnetic spectrum from 1 Hz to 1 PHz with corresponding applications radiating specific frequency bands.

A.2 Background

Any alternating electric current generates an electric and a magnetic field. If these fields are not shielded, they are radiated in form of electromagnetic waves. Depending on the type of application, this radiation is an undesired side effect (e.g., in power transmission, microwave ovens) or it is directly related to the specific purpose of the application (e.g., in wireless communication).

A.2.1 Electromagnetic Radiation

Electromagnetic waves are distinguished by their frequencies, i.e., number of oscillations per second. RF-EMFs are in the range from 3 kHz to 300 GHz, which are the frequencies the general public is most concerned about [VMZ⁺09]. The range is divided into *low-frequency* and *high-frequency* radiation fields.

Low-frequency fields. The range from 0 kHz to around 30 kHz contains low-frequency fields. The major sources emitting these fields are railway lines transmitting power at 16.7 Hz, the power transmission grid operating at 50 Hz, and electrical household appliances.

High-frequency fields. The frequencies from 30 kHz to 1 PHz count as high-frequency radiation. Most of the wireless communication and broadcasting takes place in the lower range up to several hundred GHz. The greater part of traditional radio and television broadcasting is in the band from 30 MHz to 300 MHz. These broadcast signals are generally emitted by large high-power antennas from elevated positions, generally on top of mountains with no surrounding population. The bulk of modern communication takes place in the band between 300 MHz and 6 GHz. This includes all mobile cellular system protocols like GSM, UMTS, and LTE, cordless phone technologies like DECT, as well as the heavily used 2.4 GHz band for WiFi and Bluetooth.

A.3 Electric Field Monitoring

In the following, we first describe the hardware and software requirements for commodity devices to be usable for sensing RF-EMFs. Then, we give an overview of our specific system implementation using the sensor nodes of our streetcar deployment.

A.3.1 Hardware and Software Support

A wide range of electric devices emit electromagnetic radiation. Many of these devices could be used to measure electromagnetic fields. In the following, we describe the hardware and software requirements to make these measurements feasible.

Hardware. Devices that use electromagnetic waves to receive and send data over the wireless medium, and therefore have one or multiple built-in antennas, could potentially be used to measure field strengths. Examples of such devices are widely spread in today's world with a huge amount of technical gadgets using wireless communication, such as computers, mobile phones, tablets, radios, wireless access points, and sensor boards. Typically, standard built-in antennas only support communication on specific frequency bands, and hence, also constrain RF-EMF monitoring to these bands.

Software. On the software side, it is required to have access to low-level radio information in order to read out the received signal strengths. Open-source operating systems usually provide access to this kind of information, e.g., on Linux, Android, and Arduino platforms. On closed and proprietary operating systems, such as Windows and iOS, we rely on existing APIs. For example, accessing the received signal strength is officially banned on Apple mobile devices starting from iOS 4.0.

A.3.2 System Design

We show with our system the feasibility of measuring RF-EMF exposure with commodity hardware. We extend the software on our air quality sensor nodes to measure, besides common air pollutants, the electromagnetic field strengths on different frequency bands.

GSM monitoring. We use the GSM modem on the sensor node to measure signal strengths emitted by GSM base stations. The GSM modem usually receives signals from several base stations in its vicinity. To ensure optimal connectivity, it tries to connect to the station with the highest signal strength, which becomes the serving cell. The seven strongest neighboring stations are kept in a list, and when necessary the connection

is handed over to one of these stations. For each base station in the list the modem provides information on the location area code (LAC), the unique cell ID (CID), received signal strength in dBm (RSSI), and the absolute frequency channel number (ARFCN). Requesting this information over the serial interface does not influence the data connection (i.e., ongoing data transfer is not interrupted). Hence, our application does not influence the viability of concurrent applications. The communication speed of the modem bounds the minimum measurement interval to 5 s.

WiFi monitoring. Each sensor node has an integrated MikroTik RB433 for fast and flexible data communication over WiFi for debugging and during maintenance phases. The MikroTik has built-in support to perform spectral scans fully covering the 2.4 GHz (2,182 MHz–2,549 MHz) and 5 GHz (4,790 MHz–6,085 MHz) frequency bands. The scans are intended to check the spectrum for interference from other applications to select the best channel for optimized WiFi performance when configured as access point. We use the spectral scans to monitor the exposure levels on the two WiFi bands, which are emitted by WiFi capable devices as well as non-WiFi devices emitting on the monitored frequency bands, e.g., microwave ovens, Bluetooth, and ZigBee devices.

Limitations. Sensor mobility trades off temporal coverage for an increased spatial coverage. The nodes' mobility facilitates to obtain a high spatial measurement resolution and to increase the covered area without the need of hundreds or thousands of sensor nodes. But, compared to static nodes the temporal resolution of individual locations is lower. However, observations of the spatial variation of RF-EMFs are more valuable than examinations of short-term temporal changes.

Additionally, with our sensor nodes we are not able to cover all frequency bands, which are normally sampled by personal exposure meters. However, we cover GSM and WiFi frequency bands, which are extensively used for mobile communication and are large contributors to the total field strength exposure in urban environments [FMN⁺09].

A.4 Calibration and Validation

The raw signal strength measurements from commodity hardware require calibration. This is due to unknown antenna characteristics (e.g., gain) and system losses. Moreover, it is also unclear where in the receiver amplifier chain the power measurements are made.

A.4.1 General Approach

We read out the modems' power measurements P_{dBm} in dBm and convert the readings to Watt:

$$P_W = 10^{\frac{P_{\text{dBm}}}{10}} \cdot 0.001. \quad (\text{A.1})$$

The electric field strength E has a square root dependency on the received power P_W .¹ We use the method of least squares to choose the calibration parameters c_1 and c_2 for every frequency band such that the sum of squared differences between the square root of the received powers $\sqrt{P_W}$ and the reference measurements E are minimized. Our experiments show that a linear calibration curve nicely fits the raw data to the reference measurements:

$$E = c_1 + c_2 \cdot \sqrt{P_W}, \quad (\text{A.2})$$

We perform the reference measurements with the personal exposure meter EME Spy 120 [EME14]. This exposure meter model is used in many medical [FMB⁺10] and exposure assessment studies [FMN⁺09, JFR⁺10], as well as by governmental institutions for compliance measurements [Off14]. Hence, the device is well suited for the calibration and validation of our sensor nodes.

The EME Spy 120 is able to monitor the field strength E (V/m) on 12 predefined frequency bands from FM radio (88 MHz) to WiFi (2.4 GHz) and has a lower detection limit of 0.05 V/m. Besides WiFi on 5 GHz, all other bands monitored by our sensor nodes are covered. We use the minimum measurement interval of 5 s and average the measurements to conform with the intervals of the sensor node.

A.4.2 Calibration and Validation

For the calibration procedure we simultaneously collect measurements from the EME Spy 120 and one of our sensor nodes. The two devices are placed side by side. The sensor node performs one GSM and one WiFi scan every 30 s. In order to have a variety of field strengths for calibration, we move the devices from less exposed regions (e.g., surrounded by concrete walls) to more exposed areas (e.g., close to a transmitting access point) over a time period of 35 min to 70 min. The calibrated measurements nicely fit the reference measurements on all three calibrated frequency bands, as depicted in Figure A.3.

We validate the calibration by evaluating the average absolute error between the reference measurements and the calibrated sensor readings for the different monitored frequency bands, as summarized in Table A.1.

¹For GSM, we sum up the received powers from the serving base station and from all base stations in the neighbor list to get the total received power.

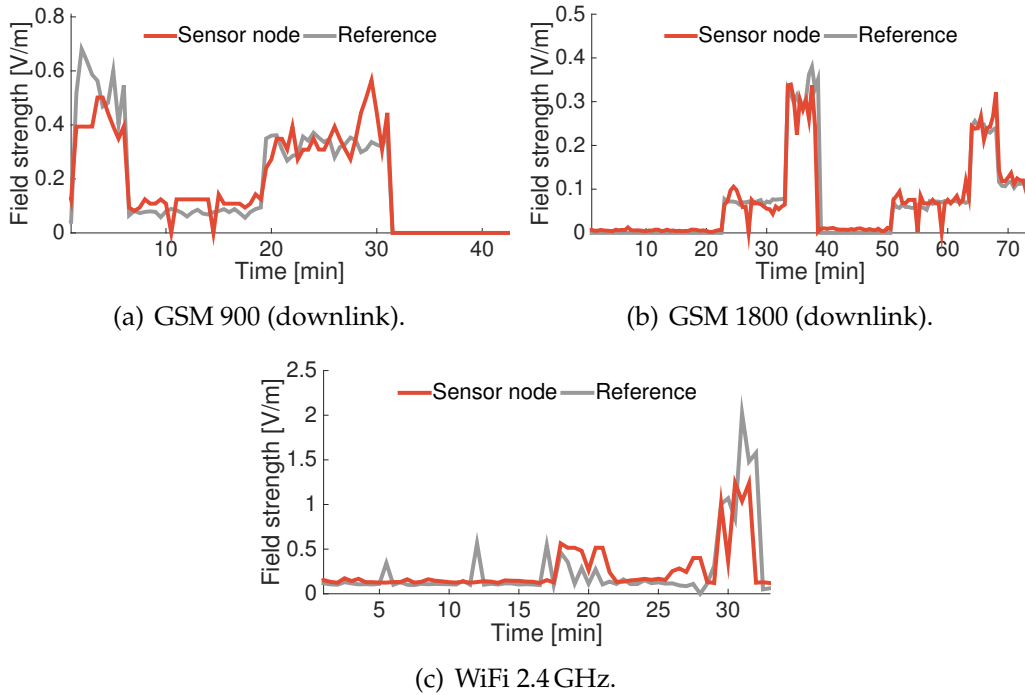


Figure A.3: Calibrated measurements from the sensor node nicely fit the reference measurements from the personal exposure meter EME Spy 120 (reference).

Frequency band	Calibration error [V/m]	Measurement range [V/m]	Pearson correlation
GSM 900	0.05	0–0.7	0.90
GSM 1800	0.02	0–0.4	0.93
WiFi 2.4 GHz	0.14	0–2.1	0.67

Table A.1: The calibration errors amount to only a few percent compared to the measurement ranges. There is a strong correlation between the measurements.

The average errors of the sensor node are 0.05 V/m and 0.02 V/m on the 900 MHz and 1800 MHz GSM bands, respectively. The latter is even below the detection limit of the reference device at 0.05 V/m. On the 2.4 GHz WiFi band we receive a slightly higher average error of 0.14 V/m. We compare the average measurement errors to the observed measurement range on the respective frequency bands, in order to get an impression of the calibration accuracy. The average measurement errors amount to only a few percent when compared to these measurement ranges, and are very low when compared to the precautionary exposure limits of 4–6 V/m [Off14]. Furthermore, all calibrated measurements show a strong correlation (Pearson $r \geq 0.6$) when compared to the reference measurements, as detailed in Table A.1.

City, Country (Period of measurement)	GSM [V/m]	WiFi [V/m]
Basel, Switzerland (2007–2008) [FMN ⁺ 09]	0.17	0.03
Ghent and Brussels, Belgium (2007–2008) [JV ⁺ 08]	0.35	0.0
Ljubljana, Slovenia (2008) [VTG09]	0.22	0.02
Utrecht, Netherland (2008) [BPK ⁺ 08]	0.36	0.0
Zurich, Switzerland (2012–2013)	0.24	0.22

Table A.2: Mean field strengths on the GSM (downlink) and 2.4 GHz WiFi bands in Zurich and five other European cities.

A.5 Measurement Campaign

We extended the software running on five sensor nodes to measure RF-EMFs emitted by GSM and WiFi communication. The nodes monitor the GSM signal strength every 15 s and transmit the data to the GSN back-end server as soon as a data connection is available. The cell towers in Zurich are run by three mobile network providers, namely Swisscom, Orange, and Sunrise. We alternate monitoring between these three operators every 10 min. We perform a spectral scan of the 2.4 GHz WiFi band every 30 s for a duration of 5 s.

A.5.1 Preliminary Findings

We collected with the streetcar deployment over the course of 4 months more than 4 million individual field strength measurements distributed over an urban area of 100 km². This is by far more than what governmental control units are able to perform manually, i.e., in total 14 measurement events at seven sensitive locations in Zurich in the period 2009–2012. In the following, we present preliminary findings based on our data set.

Transferability of calibration parameters. We use the same calibration parameters for all five identically constructed sensor nodes. We assess the transferability of the calibration parameters by comparing the sensor nodes' measurements whenever they are in each other's spatial vicinity. The average measurement deviation among the sensor nodes is small: 0.03 V/m on both GSM bands and 0.08 V/m on the 2.4 GHz WiFi band. Hence, we conclude that sensor nodes of identical construction can use the same calibration parameters without inducing a noticeable measurement inaccuracy.

International comparison. We compare the collected field strength data set from Zurich to measurements performed by others with traditional instruments in different urban cities across Europe [JFR⁺10], namely

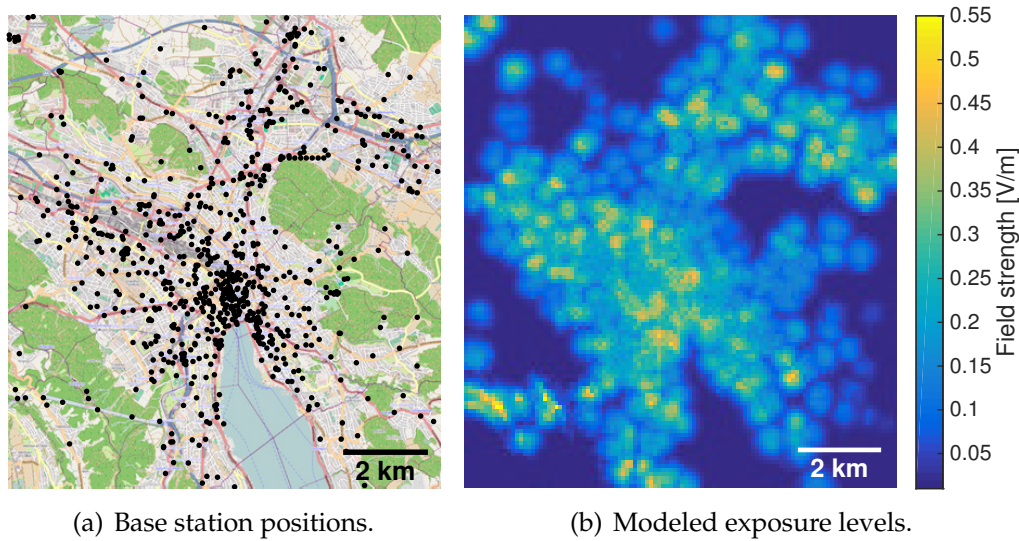


Figure A.4: Base station positions and modeled GSM exposure map of Zurich (Switzerland) with a spatial resolution of 100 m.

Basel (Switzerland), Ghent and Brussels (Belgium), Ljubljana (Slovenia), and Utrecht (Netherlands), as detailed in Table A.2. The field strength exposure on the GSM (downlink) bands in Zurich are similar to those of other cities in Europe. On the 2.4 GHz WiFi band we observe 9x higher exposure levels than what is reported from other cities. We conjecture that this is due to the rise of private and public WiFi access points over the last years. Besides our campaign, all summarized exposure assessments were carried out more than five years ago in 2007 and 2008.

A.5.2 Developing an Exposure Map

We use our spatially resolved measurements to derive high-resolution exposure maps, which can be used by the public and in epidemiological studies to assess the exposure levels at a high spatial scale. For this we apply land-use regression modeling based on General Additive Models (GAMs) [HT90], as described in Section 5.4.

We demonstrate the process by using the GSM data set to create a map of the GSM downlink signal strengths. To get a map with 100 m resolution, we first project all measurements on a grid with 100 m x 100 m cells. We develop a GAM model to calculate for each cell with measurements the dependency on the explanatory variables, which are: (i) number of base stations per cell with low, medium, and high transmission powers (classified by the governmental approval authority), (ii) distances to the next cell with a low, medium, and high transmission base station, (iii) topography of the region (i.e., elevation, slope, and aspect), (iv) total

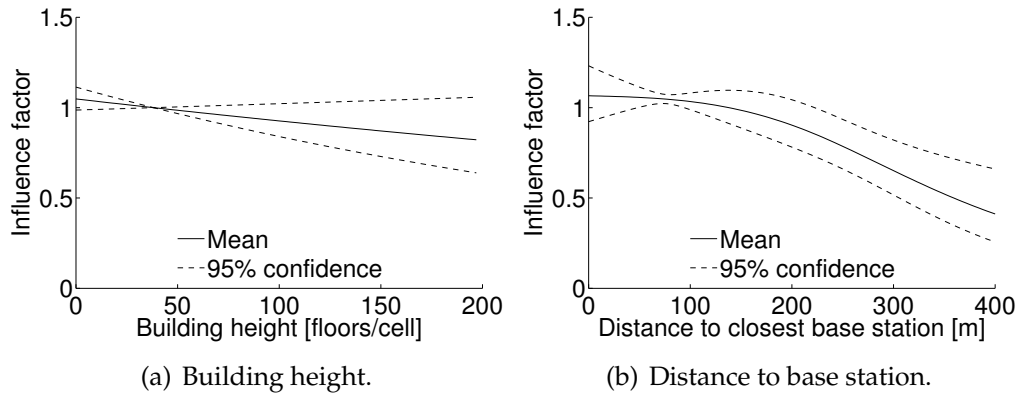


Figure A.5: Increasing building heights and distance to GSM base stations result in a decreased field strength exposure.

number of floor levels per cell, and (*v*) building density.

Figure A.4(a) depicts the GSM base station positions in the measurement region. The modeled field strengths emitted by GSM base stations are illustrated in Figure A.4(b). The modeled field strengths are mainly driven by the building height and the distance to the closest base station, as depicted in Figure A.5. As expected, both explanatory variables show a negative correlation with the measured field strength (see decreasing influence factor). We assess the model accuracy by looking at the root-mean-square error, which quantifies the differences between the modeled and measured field strengths, and the adjusted coefficient of determination (R^2), which indicates from 0 to 1 how well the modeled field strengths fit the measurements ($R^2 = 1$ denotes a perfect fit). The model has a good accuracy with a root-mean-square error of 0.12 V/m and a R^2 of 0.37. This is low but in the range of recent air quality land-use regression models having R^2 values between 0.17 and 0.9 [HBdH⁺08].

A.6 Related Work

Measurement devices. Traditionally, RF-EMFs are either monitored with personal exposure meters [Mas14, EME14] or professional spectrum analyzers [Nar14], as depicted in Figure A.6. Personal exposure meters monitor signal strengths on a set of predefined frequency bands. Professional spectrum analyzers, such as the SRM-3006 (Narda), are able to measure RF-EMF exposure on the whole spectrum. Both device types are expensive (e.g., the ESM-140 is around 4,000 USD), bulky, and tedious to use. The mobile sensor nodes achieve city-wide coverage with high spatial resolution inducing low expenditure of human labor.



Figure A.6: State-of-the-art RF-EMF exposure meters for personal (a) and professional (b) use.

Ioriatti et al. [IMV⁺09] and Viani et al. [VDO⁺11] present specific designed sensor nodes to monitor electromagnetic radiation on frequency bands from 85 MHz to 2 GHz. The detection threshold is at 0.5 V/m (compared to 0.05 V/m of our reference device), which is considerably above the mean signal strengths measured in Zurich and various European cities (see Table A.2).

Rayanchu et al. [RPB11] propose Airshark, a system that uses commodity WiFi cards to detect and classify the prevalence of non-WiFi RF devices. The spectral scans of the sensor nodes also capture the signal strengths emitted by non-WiFi devices operating in the WiFi frequency range. The Airshark approach could be used to apportion the total measured field strength to the different non-WiFi emitters.

Several Android applications exist that monitor signal strengths emitted by cellular base stations and WiFi access points, such as OpenSignal, G-MoN, and RF Signal Tracker. All these applications simply log the received signal strength in dBm. However, we require the exposure levels in V/m to compare the measurements to precautionary exposure limits and to data collected with traditional instruments. Additionally, the accuracy of these measurements is unknown.

Models. Many medical studies avail exposure models to assess radiation levels [FMBF⁺12]. Simple proxies, such as distance to closest cell

tower [ETB⁺10] or high-voltage power line [DVKS05], are inaccurate in urban environments with differing propagation conditions. More complex propagation models depend on a rich set of input data, such as antenna downtilt angles and detailed geometric building models, but can compute exposure maps with up to 5 m resolution [BFT⁺10]. We strike a balance between model complexity and spatial resolution by developing a land-use regression model to build a field strength exposure map with a spatial resolution of 100 m.

Localization and coverage maps. Signal strengths measurements are widely used for various localization techniques, e.g., [YWL12, CLLP12, ZZL⁺12, XFM⁺13]. For example, Yang et al. [YWL12] collect WiFi, and Chen et al. [CLLP12] WiFi and FM signal strengths for accurate indoor localization. Zhou et al. [ZZL⁺12] combine GSM signal strengths from several cell towers for an indoor/outdoor detection service. All these techniques rely on the relative difference between measured signal strengths as location indicator. In turn, for meaningful RF-EMF exposure monitoring we depend on accurate absolute measurements. Cellular coverage maps, e.g., from the OpenSignal project, solely provide information on the signal quality (weak to strong) but not on the absolute measured values.²

A.7 Conclusions

Radio-frequency electromagnetic fields are omnipresent in everyday life emitted by many modern applications, such as radio broadcasting and wireless communication. Up to date, electromagnetic fields are monitored with bulky and expensive devices, which have various drawbacks limiting their applicability in large-scale and long-term studies. In this appendix, we show that accurate electromagnetic field monitoring is feasible with commodity hardware. Our measurement campaign in Zurich, Switzerland, shows that radiation emitted by cellular base stations is at par with exposure levels measured in other European cities 4 to 5 years ago. However, on the 2.4 GHz WiFi band we reveal a 9-fold exposure increase over the same time period. Further, we use the collected data to develop a high-resolution GSM exposure map.

²<http://www.opensignal.com>

Bibliography

- [ABB⁺98] A. Ahlbom, U. Bergqvist, J. Bernhardt, J. Cesarini, M. Grandolfo, M. Hietanen, et al. Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). *Health Physics*, 74(4), 1998.
- [AHS06] K. Aberer, M. Hauswirth, and A. Salehi. A middleware for fast and flexible sensor network deployment. In *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB)*, pages 1199–1202, 2006.
- [Alp14a] Alphasense. CO-B4 4-Electrode carbon monoxide sensor (datasheet). <http://www.alphasense.com/WEB1213/wp-content/uploads/2014/10/COB4.pdf>, 2014.
- [Alp14b] Alphasense. NO2-B4 4-Electrode nitrogen dioxide sensor (datasheet). <http://www.alphasense.com/WEB1213/wp-content/uploads/2013/11/NO2B4.pdf>, 2014.
- [ALPR12] J. Aslam, S. Lim, X. Pan, and D. Rus. City-scale traffic estimation from a roving sensor network. In *Proceedings of the 10th International Conference on Embedded Network Sensor Systems (SenSys)*, pages 141–154. ACM, 2012.
- [AR13] D. Allemann and M. Raubal. Towards health-optimal routing in urban areas. In *Proceedings of the 2nd International Workshop on the Use of GIS in Public Health (HealthGIS)*, pages 56–59. ACM, 2013.
- [ASC⁺10] K. Aberer, S. Sathe, D. Chakraborty, A. Martinoli, G. Barrenetxea, B. Faltings, and L. Thiele. OpenSense: open community driven sensing of environment. In *In the Proceedings of the 1st International Workshop on GeoStreaming (IWGS)*, pages 39–42. ACM, 2010.
- [BBF⁺11] J. Beutel, B. Buchli, F. Ferrari, M. Keller, M. Zimmerling, and L. Thiele. X-Sense: Sensing in extreme environments. In *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2011, pages 1–6. IEEE, 2011.
- [BEH⁺06] J. A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. *Center for Embedded Network Sensing (CENS)*, 2006.

- [BFT⁺10] A. Bürgi, P. Frei, G. Theis, E. Mohler, C. Braun-Fahrländer, J. Fröhlich, et al. A model for radiofrequency electromagnetic field predictions at outdoor and indoor locations in the context of epidemiological research. *Bioelectromagnetics*, 31(3), 2010.
- [BGH⁺09] J. Beutel, S. Gruber, A. Hasler, R. Lim, A. Meier, C. Plessl, I. Talzi, L. Thiele, C. Tschudin, M. Woehrle, and M. Yuceel. Permadaq: A scientific instrument for precision sensing and data recovery in environmental extremes. In *Proceedings of the 8th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 265–276. ACM/IEEE, 2009.
- [BHK⁺11] I. Barmpadimos, C. Hueglin, J. Keller, S. Henne, and A. Prévôt. Influence of meteorology on PM₁₀ trends and variability in Switzerland from 1991 to 2008. *Atmospheric Chemistry and Physics*, 11(4):1813–1835, 2011.
- [BM98] A. Boscolo and C. Mangiavacchi. Pollution dynamic monitoring system for city air quality control. In *Proceedings of the 15th International Instrumentation and Measurement Technology Conference (IMTC)*, 1998.
- [BMEP03] V. Bychkovskiy, S. Megerian, D. Estrin, and M. Potkonjak. A collaborative approach to in-place sensor calibration. In *Proceedings of the 2nd International Conference on Information Processing in Sensor Networks (IPSN)*, pages 301–316. ACM/IEEE, 2003.
- [BN07] L. Balzano and R. Nowak. Blind calibration of sensor networks. In *Proceedings of the 6th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 79–88. ACM/IEEE, 2007.
- [BPK⁺08] J. F. Bolte, M. J. Pruppers, J. Kamer, G. Van der Zande, C. M. Schipper, S. Fleurke, T. Kluwer, I. Van Kamp, and J. Kromhout. The dutch exposimeter study: Developing an activity exposure matrix. *Epidemiology*, 19(6):78–79, 2008.
- [BRA⁺12] X. Basagaña, M. Rivera, I. Aguilera, D. Agis, L. Bouso, R. Elosua, M. Foraster, A. de Nazelle, M. Nieuwenhuijsen, J. Vila, and N. Kuenzli. Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmospheric Environment*, 54:634–642, 2012.
- [BYL⁺11] B. Buchli, M. Yuceel, R. Lim, T. Gsell, and J. Beutel. Demo abstract: Feature-rich platform for WSN design space exploration. In *Proceedings of the 10th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 115–116. ACM/IEEE, 2011.

-
- [CCD⁺11] M. Ceriotti, M. Corrà, L. D’Orazio, R. Doriguzzi, D. Facchin, S. Guna, G. P. Jesi, R. Lo Cigno, L. Mottola, A. L. Murphy, et al. Is there light at the ends of the tunnel? wireless sensor networks for adaptive lighting in road tunnels. In *Proceedings of the 10th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 187–198. ACM/IEEE, 2011.
- [CCG07] A. Carullo, S. Corbellini, and S. Grassini. A remotely controlled calibrator for chemical pollutant measuring-units. *IEEE Transactions on Instrumentation and Measurement*, 56(4):1212–1218, 2007.
- [CH04] J. Chang and S. Hanna. Air quality model performance evaluation. *Meteorology and Atmospheric Physics*, 87(1-3):167–196, 2004.
- [CH10] A. Carroll and G. Heiser. An analysis of power consumption in a smartphone. In *Proceedings of the USENIX Annual Technical Conference (ATC)*, pages 271–285. USENIX, 2010.
- [CKCH09] S. Choi, N. Kim, H. Cha, and R. Ha. Micro sensor node for air pollutant monitoring: Hardware and software issues. *Sensors*, 9(10):7970–7987, 2009.
- [CLBR10] O. Chipara, C. Lu, T. C. Bailey, and G.-C. Roman. Reliable clinical monitoring using wireless sensor networks: experiences in a step-down hospital unit. In *Proceedings of the 8th International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 155–168. ACM, 2010.
- [CLCH⁺11] S. Clifford, S. Low C., T. Hussein, K. Mengersen, and L. Morawska. Using the generalised additive model to model the particle number count of ultrafine particles. *Atmospheric Environment*, 45(32):5934–5945, 2011.
- [CLL⁺14] 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 International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 251–265. ACM, 2014.
- [CLLP12] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha. FM-based indoor localization. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 169–182. ACM, 2012.
- [CMC⁺01] 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 & Actuators B: Chemical*, 76(1):336–342, 2001.

- [CMP⁺09] M. Ceriotti, L. Mottola, G. P. Picco, A. L. Murphy, S. Guna, M. Corra, M. Pozzi, D. Zonta, and P. Zanon. Monitoring heritage buildings with wireless sensor networks: The Torre Aquila deployment. In *Proceedings of the 2009 International Conference on Information Processing in Sensor Networks (IPSN)*, pages 277–288. ACM/IEEE, 2009.
- [CP03] R. Cheng and S. Prabhakar. Managing uncertainty in sensor database. *ACM SIGMOD Record*, 32(4):41–46, 2003.
- [CRSC06] R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. *Measurement error in nonlinear models: a modern perspective*. Chapman & Hall/CRC, 2006.
- [CZZ⁺13] C. Chen, D. Zhang, Z.-H. Zhou, N. Li, T. Atmaca, and S. Li. B-planner: Night bus route planning using large-scale taxi gps traces. In *Proceedings of the 11th International Conference on Pervasive Computing and Communications (PerCom)*, pages 225–233. IEEE, 2013.
- [DAK⁺09] P. Dutta, P. M. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff. Demo abstract: Common sense: participatory urban sensing using a network of handheld air quality monitors. In *Proceedings of the 7th International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 349–350. ACM, 2009.
- [Dem44] W. E. Deming. *Statistical adjustment of data*. New York, 1944.
- [DHBNH10] J. J. De Hartog, H. Boogaard, H. Nijland, and G. Hoek. Do the health benefits of cycling outweigh the risks? *Environmental Health Perspectives*, 118(8):1109–1116, 2010.
- [DSL⁺13] 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 Workshop on Urban Computing (UrbComp)*. ACM, 2013.
- [DVKS05] G. Draper, T. Vincent, M. Kroll, and J. Swanson. Childhood cancer in relation to distance from high voltage power lines in england and wales: a case-control study. *BMJ*, 330(7503), 2005.
- [DVMP⁺08] S. De Vito, E. Massera, M. Piga, L. Martinotto, and G. Di Francia. On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors & Actuators B: Chemical*, 129(2):750–757, 2008.
- [EME14] EME SPY 120. Personal exposure meter for high-frequency fields. <http://www.satimo.com>, 2014.

- [EN03a] E. Elnahrawy and B. Nath. Cleaning and querying noisy sensors. In *Proceedings of the 2nd International Conference on Wireless Sensor Networks and Applications (WSNA)*, pages 78–87. ACM, 2003.
- [EN03b] E. Elnahrawy and B. Nath. Online data cleaning in wireless sensor networks. In *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 294–295. ACM, 2003.
- [ETB⁺10] P. Elliott, M. B. Toledano, J. Bennett, L. Beale, K. De Hoogh, N. Best, and D. Briggs. Mobile phone base stations and early childhood cancers: case-control study. *BMJ*, 340, 2010.
- [ETH14] ETH Zurich. PermaSense and X-Sense website. <http://www.permasense.ch>, 2014.
- [FB92] E. D. Feigelson and G. J. Babu. Linear regression in astronomy. II. *The Astrophysical Journal*, 397:55–67, 1992.
- [FBG⁺02] T. R. Fears, C. C. Bird, D. Guerry, R. W. Sagebiel, M. H. Gail, D. E. Elder, A. Halpern, E. A. Holly, P. Hartge, and M. A. Tucker. Average midrange ultraviolet radiation flux and time outdoors predict melanoma risk. *Cancer Research*, 62(14):3992–3996, 2002.
- [Fed09] Federal Office for the Environment. Measurement results from the Swiss National Air Pollution Monitoring Network (NABEL) for the year 2009. EMPA, 2009.
- [Fed13] Federal Office for the Environment. Measurement results from the Swiss National Air Pollution Monitoring Network (NABEL) for the year 2012. EMPA, 2013.
- [FHSB11] M. Fierz, C. Houle, P. Steigmeier, and H. Burtscher. Design, calibration, and field performance of a miniature diffusion size classifier. *Aerosol Science and Technology*, 45(1):1–10, 2011.
- [Fit06] P. Fitzpatrick. *Advanced calculus*, volume 5. American Mathematical Soc., 2006.
- [FKS⁺11] T. Fahrni, M. Kuhn, P. Sommer, R. Wattenhofer, and S. Welten. Sundroid: Solar radiation awareness with smartphones. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp)*, pages 365–374. ACM, 2011.
- [FMB⁺10] P. Frei, E. Mohler, A. Bürgi, J. Fröhlich, G. Neubauer, C. Braun-Fahrländer, and M. Röösli. Classification of personal exposure to radio frequency electromagnetic fields (RF-EMF) for epidemiological research: Evaluation of different exposure assessment methods. *Environment International*, 36(7), 2010.

- [FMBF⁺12] P. Frei, E. Mohler, C. Braun-Fahrländer, J. Fröhlich, G. Neubauer, and M. Rössli. Cohort study on the effects of everyday life radio frequency electromagnetic field exposure on non-specific symptoms and tinnitus. *Environment International*, 38(1), 2012.
- [FMN⁺09] P. Frei, E. Mohler, G. Neubauer, G. Theis, A. Bürgi, J. Fröhlich, C. Braun-Fahrländer, J. Bolte, M. Egger, and M. Rössli. Temporal and spatial variability of personal exposure to radio frequency electromagnetic fields. *Environmental research*, 109(6):779–785, 2009.
- [FMP03] J. Feng, S. Megerian, and M. Potkonjak. Model-based calibration for sensor networks. In *Proceedings of the 2nd International Conference on Sensors*, volume 2, pages 737–742. IEEE, 2003.
- [FOC⁺11] M. Faulkner, M. Olson, R. Chandy, J. Krause, K. M. Chandy, and A. Krause. The next big one: Detecting earthquakes and other rare events from community-based sensors. In *Proceedings of the 10th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 13–24. ACM/IEEE, 2011.
- [FT00] C. Frost and S. G. Thompson. Correcting for regression dilution bias: comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society Series A*, 163(2):173–189, 2000.
- [Gem14] Gemalto. Cinterion MC75i. http://m2m.gemalto.com/tl_files/cinterion/downloads/datasheet_scalableplatformmc75itc65itc63i_175016.pdf, 2014.
- [GH09] V. C. Gungor and G. P. Hancke. Industrial wireless sensor networks: Challenges, design principles, and technical approaches. *IEEE Transactions on Industrial Electronics*, 56(10):4258–4265, 2009.
- [Gil06] J. Gillard. An historical overview of linear regression with errors in both variables. *Technical Report Cardiff University*, 2006.
- [Goo14] GoogleMaps. Route planner. <http://www.maps.google.com>, 2014.
- [GPA⁺10] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher. Greengps: a participatory sensing fuel-efficient maps application. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 151–164. ACM, 2010.
- [Gum14] Gumstix. Gumstix Verdex Pro XL6P COM. <https://store.gumstix.com/index.php/products/210/>, 2014.
- [HA04] V. J. Hodge and J. Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004.

- [HBdH⁺08] G. Hoek, R. Beelen, K. de Hoogh, D. Vienneau, J. Gulliver, P. Fischer, and D. Briggs. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, 42(33):7561–7578, 2008.
- [HBK⁺09] G. Hoek, H. Boogaard, A. Knol, J. De Hartog, P. Slottje, J. G. Ayres, P. Borm, B. Brunekreef, K. Donaldson, F. Forastiere, et al. Concentration response functions for ultrafine particles and all-cause mortality and hospital admissions: results of a European expert panel elicitation. *Environmental Science & Technology*, 44(1):476–482, 2009.
- [HBPW08] R. Honicky, E. A. Brewer, E. Paulos, and R. White. N-smarts: networked suite of mobile atmospheric real-time sensors. In *Proceedings of the 2nd Workshop on Networked Systems for Developing Regions (NSDR)*, pages 25–30. ACM, 2008.
- [HHK⁺08] O. Hertel, M. Hvidberg, M. Ketzel, L. Storm, and L. Stausgaard. A proper choice of route significantly reduces air pollution exposure—a study on bicycle and bus trips in urban streets. *Science of the Total Environment*, 389(1):58–70, 2008.
- [HNR68] P. E. Hart, N. J. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *Systems Science and Cybernetics*, 4(2):100–107, 1968.
- [HR97] M. M. Halldórsson and J. Radhakrishnan. Greed is good: Approximating independent sets in sparse and bounded-degree graphs. *Algorithmica*, 18(1):145–163, 1997.
- [HST12] D. Hasenfratz, O. Saukh, and L. Thiele. On-the-fly calibration of low-cost gas sensors. In *Proceedings of the 9th European Conference on Wireless Sensor Networks (EWSN)*, pages 228–244. Springer, 2012.
- [HT90] T. J. Hastie and R. J. Tibshirani. *Generalized additive models*, volume 43. Chapman & Hall/CRC, 1990.
- [HY00] R. M. Harrison and J. Yin. Particulate matter in the atmosphere: which particle properties are important for its effects on health? *Science of The Total Environment*, 249(1):85–101, 2000.
- [HZG11] C. Huan, L. Zhiyu, and F. Gang. Analysis of the aging characteristics of SnO₂ gas sensors. *Sensors & Actuators, B: Chemical*, 156(2):912–917, 2011.
- [IMV⁺09] L. Ioriatti, M. Martinelli, F. Viani, M. Benedetti, and A. Massa. Real-time distributed monitoring of electromagnetic pollution in urban environments. In *Proceedings of the 29th International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 5, pages 100–103. IEEE, 2009.

- [JAK⁺04] M. Jerrett, A. Arain, P. Kanaroglou, B. Beckerman, D. Potoglou, T. Sahsuvaroglu, J. Morrison, and C. Giovis. A review and evaluation of intraurban air pollution exposure models. *Exposure Science and Environmental Epidemiology*, 15(2):185–204, 2004.
- [JCCE09] L. Jackson, N. Carslaw, D. Carslaw, and K. Emmerson. Modelling trends in OH radical concentrations using generalized additive models. *Atmospheric Chemistry and Physics*, 9(6):2021–2033, 2009.
- [JFR⁺10] W. Joseph, P. Frei, M. Roösli, G. Thuróczy, P. Gajsek, T. Trcek, et al. Comparison of personal radio frequency electromagnetic field exposure in different urban areas across Europe. *Environmental Research*, 110(7), 2010.
- [JJWF13] X. Jiang, J. Jia, G. Wu, and J. Z. Fang. Low-cost personal air-quality monitor. In *Proceeding of the 11th International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 491–492. ACM, 2013.
- [JLT⁺11] Y. Jiang, K. Li, L. Tian, R. Piedrahita, X. Yun, O. Mansata, Q. Lv, R. P. Dick, M. Hannigan, and L. Shang. Maqs: A mobile sensing system for indoor air quality. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp)*, pages 493–494. ACM, 2011.
- [JVV⁺08] W. Joseph, G. Vermeeren, L. Verloock, M. M. Heredia, and L. Martens. Characterization of personal RF electromagnetic field exposure and actual absorption for the general public. *Health physics*, 95(3):317–330, 2008.
- [Kan10] E. Kanjo. Noisespys: A real-time mobile phone platform for urban noise monitoring and mapping. *Mobile Networks and Applications*, 15(4):562–574, 2010.
- [KBP06] M. Kamionka, P. Breuil, and C. Pijolat. Calibration of a multivariate gas sensing device for atmospheric pollution measurement. *Sensors & Actuators B: Chemical*, 118(1):323–327, 2006.
- [KBST12] M. Keller, J. Beutel, O. Saukh, and L. Thiele. Visualizing large sensor network data sets in space and time with Vizzly. In *Proceedings of the 7th International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp)*, pages 925–933. IEEE, 2012.
- [KEH⁺13] Y. Kim, J. Eberle, R. Hanninen, E. C. Un, and K. Aberer. Mobile observatory: an exploratory study of mobile air quality monitoring application. In *Proceedings of the 15th International Conference on Pervasive and Ubiquitous Computing (UbiComp)*, pages 733–736. ACM, 2013.

-
- [KGZ07] A. Kansal, M. Goraczko, and F. Zhao. Building a sensor network of mobile phones. In *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 547–548. ACM, 2007.
- [KH50] K. A. Kermack and J. B. S. Haldane. Organic correlation and allometry. *Biometrika*, 37:30–41, 1950.
- [KHB⁺03] H. Kruize, O. Hänninen, O. Breugelmans, E. Lebet, and M. Jantunen. Description and demonstration of the expolis simulation model: two examples of modeling population exposure to particulate matter. *Journal of Exposure Science and Environmental Epidemiology*, 13(2):87–99, 2003.
- [KHLH14] A. Kapoor, Z. Horvitz, S. Laube, and E. Horvitz. Airplanes aloft as a sensor network for wind forecasting. In *Proceedings of the 13th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 25–34. ACM/IEEE, 2014.
- [KPM11] E. Koukoumidis, L.-S. Peh, and M. R. Martonosi. Signalguru: leveraging mobile phones for collaborative traffic signal schedule advisory. In *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 127–140. ACM, 2011.
- [Kru53] W. H. Kruskal. On the uniqueness of the line of organic correlation. *Biometrics*, 9:47–58, 1953.
- [KTK⁺09] V. Khurana, C. Teo, M. Kundi, L. Hardell, and M. Carlberg. Cell phones and brain tumors: a review including the long-term epidemiologic data. *Surgical Neurology*, 72(3), 2009.
- [KWL⁺11] M. Keller, M. Woehrle, R. Lim, J. Beutel, and L. Thiele. Comparative performance analysis of the PermaDozer protocol in diverse deployments. In *Proceedings of the 6th International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp)*, pages 957–965. IEEE, 2011.
- [Lan99] R. B. Langley. Dilution of precision. *GPS World*, 10(5):52–59, 1999.
- [LB14] J. Lipor and L. Balzano. Robust blind calibration via total least squares. In *Proceedings of the 39th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4244–4248. IEEE, 2014.
- [LFS⁺12] J. J. Li, B. Faltings, O. Saukh, D. Hasenfratz, and J. Beutel. Sensing the air we breathe—the OpenSense Zurich dataset. In *Proceedings of the 26th International Conference on Artificial Intelligence (AAAI)*, 2012.

- [LHL⁺11] Y. Liu, Y. He, M. Li, J. Wang, K. Liu, L. Mo, W. Dong, Z. Yang, M. Xi, J. Zhao, et al. Does wireless sensor network scale? a measurement study on greenorbs. In *Proceedings of the 30th International Conference on Computer Communications (INFOCOM)*, pages 873–881. IEEE, 2011.
- [LLL⁺09] C.-J. M. Liang, J. Liu, L. Luo, A. Terzis, and F. Zhao. RACnet: a high-fidelity data center sensing network. In *Proceedings of the 7th International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 15–28. ACM, 2009.
- [LPK09] Y. Liu, C. J. Paciorek, and P. Koutrakis. Estimating regional spatial and temporal variability of PM_{2.5} concentrations using satellite data, meteorology, and land use information. *Environmental Health Perspectives*, 117(6):886–892, 2009.
- [LSA01] E. Limpert, W. A. Stahel, and M. Abbt. Log-normal distributions across the sciences: Keys and clues. *BioScience*, 51(5):341–352, 2001.
- [Lun14] Lunge Zurich. <http://www.lunge-zuerich.ch>, 2014.
- [LUWP11] S. Lindenthal, V. Ussyshkin, J. Wang, and M. Pokorny. Airborne LIDAR: A fully-automated self-calibration procedure. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, 3812:73–78, 2011.
- [Map14] MapQuest. Route planner. <http://www.mapquest.com>, 2014.
- [Mas14] Maschek ESM-140. Personal exposure meter for high-frequency fields. <http://www.maschek.de>, 2014.
- [MHCH12] B. Mølgaard, T. Hussein, J. Corander, and K. Hämeri. Forecasting size-fractionated particle number concentrations in the urban atmosphere. *Atmospheric Environment*, 46:155–163, 2012.
- [MHL⁺09] L. Mo, Y. He, Y. Liu, J. Zhao, S.-J. Tang, X.-Y. Li, and G. Dai. Canopy closure estimates with greenorbs: sustainable sensing in the forest. In *Proceedings of the 7th International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 99–112. ACM, 2009.
- [Mik14] MikroTik. Routerboard 433. 2014.
- [MJK⁺10] S. Mathur, T. Jin, N. Kasturirangan, J. Chandrasekaran, W. Xue, M. Gruteser, and W. Trappe. Parknet: drive-by sensing of road-side parking statistics. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pages 123–136. ACM, 2010.

- [MLB⁺12] J. Monroy, A. Lilienthal, J. Blanco, J. González-Jimenez, and M. Trincavelli. Calibration of MOX gas sensors in open sampling systems based on gaussian processes. In *Proceedings of the 11th International Conference on Sensors*, pages 1743–1746. IEEE, 2012.
- [MLCOS08] E. Miluzzo, N. D. Lane, A. T. Campbell, and R. Olfati-Saber. CaliBree: A self-calibration system for mobile sensor networks. In *Proceedings of the 4th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 314–331. IEEE, 2008.
- [MLE⁺15] S. Moltchanov, I. Levy, Y. Etzion, U. Lerner, D. M. Broday, and B. Fishbain. On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Science of The Total Environment*, 502:537–547, 2015.
- [MLF⁺08] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell. Sensing meets mobile social networks: the design, implementation and evaluation of the CenceMe application. In *Proceedings of the 6th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 337–350. ACM/IEEE, 2008.
- [MOP⁺96] D. Mage, G. Ozolins, P. Peterson, A. Webster, R. Orthofer, V. Vandeweerd, and M. Gwynne. Urban air pollution in megacities of the world. *Atmospheric Environment*, 30(5):681–686, 1996.
- [MPR08] P. Mohan, V. N. Padmanabhan, and R. Ramjee. Nericell: Rich monitoring of road and traffic conditions using mobile smartphones. In *Proceedings of the 6th International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 323–336. ACM, 2008.
- [MPSS13] C. Meurisch, K. Planz, D. Schäfer, and I. Schweizer. Noisemap: Discussing scalability in participatory sensing. In *Proceedings of 1st International Workshop on Sensing and Big Data Mining (SenseMine)*. ACM, 2013.
- [MSPD09] S. Mukhopadhyay, C. Schurgers, D. Panigrahi, and S. Dey. Model-based techniques for data reliability in wireless sensor networks. *IEEE Transactions on Mobile Computing*, 8(4):528–543, 2009.
- [MVH07] I. Markovsky and S. Van Huffel. Overview of total least-squares methods. *Signal Processing*, 87(10):2283–2302, 2007.
- [Nar14] Narda SRM-3006. Selective radiation meter for high-frequency fields. <http://www.narda-sts.de>, 2014.
- [Nat14a] National Air Pollution Monitoring Network (NABEL). <http://www.bafu.admin.ch/luft/00612/00625/index.html?lang=en>, 2014.

- [Nat14b] National Air Pollution Monitoring Network (NABEL). Air pollution data from the last 18 months. http://www.bafu.admin.ch/luft/luftbelastung/blick_zurueck/datenabfrage/index.html?lang=en, 2014.
- [Nyf01] U. Nyffeler. Das Nationale Beobachtungsnetz für Luftfremdstoffe. BUWAL, 2001.
- [Off14] Office of Waste, Water, Energy, and Air (WWEA). RF-EMF audits in Zurich. http://www.awel.zh.ch/internet/audirektion/awel/de/luft_klima_elektrosmog/elektrosmog/messungen/nis_definitionsseite/zuerich.html, 2014.
- [PPS⁺99] C. Pijolat, C. Pupier, M. Sauvan, G. Tournier, and R. Lalauze. Gas detection for automotive pollution control. *Sensors & Actuators B: Chemical*, 59(2):195–202, 1999.
- [PSLN⁺12] M. Pajic, S. Sundaram, J. Le Ny, G. J. Pappas, and R. Mangharam. Closing the loop: A simple distributed method for control over wireless networks. In *Proceedings of the 11th international conference on Information Processing in Sensor Networks (IPSN)*, pages 25–36. ACM/IEEE, 2012.
- [PTPB13] J. Peters, J. Theunis, M. V. Poppel, and P. Berghmans. Monitoring PM10 and ultrafine particles in urban environments using mobile measurements. *Aerosol and Air Quality Research*, 13:509–522, 2013.
- [PTT⁺01] P. Penttinen, K. L. Timonen, P. Tiittanen, A. Mirme, J. Ruuskanen, and J. Pekkanen. Number concentration and size of particles in urban air: effects on spirometric lung function in adult asthmatic subjects. *Environmental Health Perspectives*, 109(4):319–323, 2001.
- [RBB⁺06] N. Ramanathan, L. Balzano, M. Burt, D. Estrin, T. Harmon, C. Harvey, J. Jay, E. Kohler, S. Rothenberg, and M. Srivastava. Rapid deployment with confidence: Calibration and fault detection in environmental sensor networks. *Technical Report TR 62, Center for Embedded Networked Sensing (CENS)*, 2006.
- [RCK⁺10] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu. Earphone: an end-to-end participatory urban noise mapping system. In *Proceedings of the 9th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 105–116. ACM/IEEE, 2010.
- [RCM⁺12] U. Raza, A. Camerra, A. L. Murphy, T. Palpanas, and G. P. Picco. What does model-driven data acquisition really achieve in wireless sensor networks? In *Proceedings of the 10th International Conference on Pervasive Computing and Communications (PerCom)*, pages 85–94. IEEE, 2012.

- [RDSF⁺14] M. S. Ragettli, R. E. Ducret-Stich, M. Foraster, X. Morelli, I. Aguilera, X. Basagaña, E. Corradi, A. Ineichen, M.-Y. Tsai, N. Probst-Hensch, et al. Spatio-temporal variation of urban ultrafine particle number concentrations. *Atmospheric Environment*, 96:275–283, 2014.
- [Ric73] W. E. Ricker. Linear regressions in fishery research. *Journal of the Fisheries Research Board of Canada*, 30(3):409–434, 1973.
- [RK04] L. Reznik and V. Kreinovich. Fuzzy and probabilistic models of association information in sensor networks. In *Proceedings of the International Conference on Fuzzy Systems (FUZZ)*, volume 1, pages 185–189. IEEE, 2004.
- [RMM⁺10] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas. Emotionsense: a mobile phones based adaptive platform for experimental social psychology research. In *Proceedings of the 12th International Conference on Ubiquitous Computing (UbiComp)*, pages 281–290. ACM, 2010.
- [RPB11] S. Rayanchu, A. Patro, and S. Banerjee. Airshark: detecting non-WiFi RF devices using commodity WiFi hardware. In *Proceedings of the International Conference on Internet Measurement (SIGCOMM)*, pages 137–154. ACM, 2011.
- [RPV14] G. L. Re, D. Peri, and S. D. Vassallo. Urban air quality monitoring using vehicular sensor networks. *Advances onto the Internet of Things*, 260:311–323, 2014.
- [Sam42] P. A. Samuelson. A note on alternative regressions. *Econometrica*, 10(1):80–83, 1942.
- [SBA⁺08] J. G. Su, M. Brauer, B. Ainslie, D. Steyn, T. Larson, and M. Buzzelli. An innovative land use regression model incorporating meteorology for exposure analysis. *Science of the Total Environment*, 390(2):520–529, 2008.
- [SD10] M. Stevens and E. D’Hondt. Crowdsourcing of pollution data using smartphones. In *Proceedings of the 1st Workshop on Ubiquitous Crowdsourcing*, 2010.
- [Sen68] P. K. Sen. Estimates of the regression coefficient based on kendall’s tau. *Journal of the American Statistical Association*, 63(324):1379–1389, 1968.
- [Sen14] Sensirion. SHT10 humidity and temperature sensor. http://www.sensirion.com/fileadmin/user_upload/customers/sensirion/Dokumente/Humidity/Sensirion_Humidity_SHT1x_Datasheet_V5.pdf, 2014.

- [SGG10] A. B. Sharma, L. Golubchik, and R. Govindan. Sensor faults: Detection methods and prevalence in real-world datasets. *ACM Transactions on Sensor Networks (TOSN)*, 6(3):23, 2010.
- [SGX14] SGX Sensortech. MiCS-OZ-47 ozone sensor (datasheet). http://sgx.cdistore.com/datasheets/e2v/a1a-mics-oz-47_1_s1%20081231.pdf, 2014.
- [SHWT13] O. Saukh, D. Hasenfratz, C. Walser, and L. Thiele. On rendezvous in mobile sensing networks. In *Proceedings of the 5th International Workshop on Real-World Wireless Sensor Networks (RealWSN)*, pages 29–42. Springer, 2013.
- [SK14] M. H. Sharker and H. A. Karimi. Computing least air pollution exposure routes. *International Journal of Geographical Information Science*, 28(2):343–362, 2014.
- [SLMJ07] B. Sheng, Q. Li, W. Mao, and W. Jin. Outlier detection in sensor networks. In *Proceedings of the 8th International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pages 219–228. ACM, 2007.
- [Smi09] R. J. Smith. Use and misuse of the reduced major axis for line-fitting. *American Journal of Physical Anthropology*, 140(3):476–486, 2009.
- [SP96] G. D. Smith and A. N. Phillips. Inflation in epidemiology: The proof and measurement of association between two things revisited. *British Medical Journal*, 312(7047):1659–1661, 6 1996.
- [SW02] G. Shaddick and J. Wakefield. Modelling daily multivariate pollutant data at multiple sites. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(3):351–372, 2002.
- [SZT⁺14] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu. Inferring gas consumption and pollution emission of vehicles throughout a city. In *Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pages 1027–1036. ACM, 2014.
- [Tax14] Taxi & Limousine Commission, New York City. The 2014 taxicab fact book. http://www.nyc.gov/html/tlc/downloads/pdf/2014_taxicab_fact_book.pdf, 2014.
- [TBGE10] A. Thiagarajan, J. Biagioni, T. Gerlich, and J. Eriksson. Cooperative transit tracking using smart-phones. In *Proceedings of the 8th International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 85–98. ACM, 2010.
- [The50] H. Theil. A rank-invariant method of linear and polynomial regression analysis I, II, III. In *Proceedings of Royal Netherlands*

- Academy of Sciences*, volume 53, pages 386–392, 521–525, 1397–1412, 1950.
- [Tom14] TomTom. Route planner. <http://www.routes.tomtom.com>, 2014.
- [TPS⁺05] G. Tolle, J. Polastre, R. Szewczyk, D. Culler, N. Turner, K. Tu, S. Burgess, T. Dawson, P. Buonadonna, D. Gay, et al. A macroscope in the redwoods. In *Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 51–63. ACM, 2005.
- [TLLSC11] T. Tsung-Te Lai, C.-Y. Lin, Y.-Y. Su, and H.-H. Chu. BikeTrack: Tracking stolen bikes through everyday mobile phones and participatory sensing. In *Proceedings of the 2nd International Workshop on Sensing Applications on Mobile Phones (PhoneSense)*. ACM, 2011.
- [TXY⁺13] R. Tan, G. Xing, Z. Yuan, X. Liu, and J. Yao. System-level calibration for data fusion in wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 9(3):28, 2013.
- [TYIM05] W. Tsujita, A. Yoshino, H. Ishida, and T. Moriizumi. Gas sensor network for air-pollution monitoring. *Sensors & Actuators B: Chemical*, 110(2):304–311, 2005.
- [Ub14] U-blox. EVK-6P evaluation kit. http://www.u-blox.com/images/downloads/Product_Docs/EVK-6_ProductSummary_%28GPS.G6-EK-10031%29.pdf, 2014.
- [UNO92] UNO. Rio declaration on environment and development. In *United Nations Conference on Environment and Development*, 1992.
- [VDO⁺11] F. Viani, M. Donelli, G. Oliveri, A. Massa, and D. Trincherò. A WSN-based system for real-time electromagnetic monitoring. In *Proceedings of the International Symposium on Antennas and Propagation (APSURSI)*, pages 3129–3132. IEEE, 2011.
- [VFPGF03] S. Vardoulakis, B. E. Fisher, K. Pericleous, and N. Gonzalez-Flesca. Modelling air quality in street canyons: a review. *Atmospheric Environment*, 37(2):155–182, 2003.
- [VMZ⁺09] P. Vecchia, R. Matthes, G. Ziegelberger, J. Lin, R. Saunders, and A. Swerdlow. Exposure to high frequency electromagnetic fields, biological effects and health consequences (100 kHz–300 GHz). *International Commission on Non-Ionizing Radiation Protection (ICNIRP)*, 2009.
- [VNKL08] P. Völgyesi, A. Nadas, X. Koutsoukos, and Á. Lédeczi. Demo abstract: Air quality monitoring with SensorMap. In *Proceedings*

- of the 7th International Conference on Information Processing in Sensor Networks (IPSN), pages 529–530. ACM/IEEE, 2008.
- [VTG09] B. Valic, T. Trcek, and P. Gajsek. Personal exposure to high frequency electromagnetic fields in Slovenia. In *Joint Meeting of the Bioelectromagnetics Society and the European BioElectromagnetics Association*, 2009.
- [WC02] K. Whitehouse and D. Culler. Calibration as parameter estimation in sensor networks. In *Proceedings of the 1st International Workshop on Wireless Sensor Networks and Applications (WSNA)*, pages 59–67. ACM, 2002.
- [WHO14] WHO. 7 million premature deaths annually linked to air pollution. <http://www.who.int/mediacentre/news/releases/2014/air-pollution>, 2014.
- [Woo41] E. B. Woolley. The method of minimized areas as a basis for correlation analysis. *Econometrica*, 9(1):38–62, 1941.
- [XBP⁺12] Y. Xiang, L. Bai, R. Piedrahita, R. P. Dick, Q. Lv, M. Hannigan, and L. Shang. Collaborative calibration and sensor placement for mobile sensor networks. In *Proceedings of the 11th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 73–84. ACM/IEEE, 2012.
- [XFM⁺13] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An. SCPL: Indoor device-free multi-subject counting and localization using radio signal strength. In *Proceedings of the 12th international conference on Information Processing in Sensor Networks (IPSN)*, pages 79–90. ACM/IEEE, 2013.
- [XRC⁺04] N. Xu, S. Rangwala, K. K. Chintalapudi, D. Ganesan, A. Broad, R. Govindan, and D. Estrin. A wireless sensor network for structural monitoring. In *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems (SenSys)*, pages 13–24. ACM, 2004.
- [YWL12] Z. Yang, C. Wu, and Y. Liu. Locating in fingerprint space: wireless indoor localization with little human intervention. In *Proceedings of the 18th International Conference on Mobile Computing and Networking (MobiCom)*, pages 269–280. ACM, 2012.
- [ZBP⁺12] P. Zappi, E. Bales, J. H. Park, W. Griswold, and T. Š. Rosing. The citisense air quality monitoring mobile sensor node. In *Mobile Sensing: From Smartphones and Wearables to Big Data*, 2012.
- [ZGT⁺14] B. Zou, Y. Guo, Y. Tang, S. Xu, and Q. Weng. Remote sensing detection of the spatial pattern of urban air pollution in los angeles.

- In *In Proceedings of the 3rd International Workshop on Earth Observation and Remote Sensing Applications (EORSA)*, pages 251–255. IEEE, 2014.
- [ZHKS02] Y. Zhu, W. C. Hinds, S. Kim, and C. Sioutas. Concentration and size distribution of ultrafine particles near a major highway. *Journal of the Air & Waste Management Association*, 52(9):1032–1042, 2002.
- [ZLH13] Y. Zheng, F. Liu, and H.-P. Hsie. U-Air: When urban air quality inference meets big data. In *Proceedings of the 19th Int. Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pages 1436–1444. ACM, 2013.
- [ZLW⁺14] Y. Zheng, T. Liu, Y. Wang, Y. Zhu, and E. Chang. Diagnosing New York City’s noises with ubiquitous data. In *Proceedings of the 16th International Conference on Ubiquitous Computing (UbiComp)*. ACM, 2014.
- [ZMH10] Y. Zhang, N. Meratnia, and P. Havinga. Outlier detection techniques for wireless sensor networks: A survey. *IEEE Communications Surveys & Tutorials*, 12(2):159–170, 2010.
- [ZZL⁺12] P. Zhou, Y. Zheng, Z. Li, M. Li, and G. Shen. IODetector: A generic service for indoor outdoor detection. In *Proceedings of the 10th International Conference on Embedded Network Sensor Systems (SenSys)*, pages 113–126. ACM, 2012.

List of Publications

The following list includes publications that form the basis of this thesis. The corresponding chapters are indicated in parentheses.

D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele. **Participatory air pollution monitoring using smartphones.** In *Proceedings of the 2nd International Workshop on Mobile Sensing (in conjunction with ACM/IEEE IPSN)*. Beijing, China, April 2012. (Chapter 2)

J. J. Li, B. Faltings, O. Saukh, D. Hasenfratz, and J. Beutel. **Sensing the air we breathe – the OpenSense Zurich dataset.** In *Proceedings of the 26th International Conference on Artificial Intelligence (AAAI)*. Toronto, Canada, July 2012. (Chapter 2)

O. Saukh*, D. Hasenfratz*, and L. Thiele. **Reducing multi-hop calibration errors in mobile sensor networks.** In *submission to the 14th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. Seattle, USA, April 2015. (Chapter 3)

*These authors contributed equally to this work.

D. Hasenfratz, O. Saukh, and L. Thiele. **Model-driven accuracy bounds for noisy sensor readings.** In *Proceedings of the 9th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS)*. Massachusetts, USA, May 2013. (Chapter 4)

D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, and L. Thiele. **Poster abstract: Revealing the limits of spatio-temporal high-resolution pollution maps.** In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys)*. Rome, Italy, November 2013. (Chapter 5)

D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, and L. Thiele. **Pushing the spatio-temporal resolution limit of urban air pollution maps.** In *Proceedings of the 12th IEEE International Conference on Pervasive Computing and Communications (PerCom)*. Budapest, Hungary, March 2014. (*Best paper award*. Chapter 5)

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.** In *Journal of Pervasive and Mobile Computing*. Elsevier. (Chapter 5 and Chapter 6)

D. Hasenfratz, S. Sturzenegger, O. Saukh, and L. Thiele. **Spatially resolved monitoring of radio-frequency electromagnetic fields.** In *Proceedings of the 1st International Workshop on Sensing and Big Data Mining (in conjunction with ACM SenSys)*. Rome, Italy, November 2013. (App. A)

The following list includes publications that were written during the PhD studies, yet are not part of this thesis.

O. Saukh, D. Hasenfratz, A. Noori, T. Ulrich, and L. Thiele. **Demo abstract: Route selection of mobile sensors for air quality monitoring.** In *Proceedings of the 9th European Conference on Wireless Sensor Networks (EWSN)*. Trento, Italy, February 2012.

D. Hasenfratz, O. Saukh, and L. Thiele. **On-the-fly calibration of low-cost gas sensors.** In *Proceedings of the 9th European Conference on Wireless Sensor Networks (EWSN)*. Trento, Italy, February 2012.

O. Saukh, D. Hasenfratz, A. Noori, T. Ulrich, and L. Thiele. **Route selection for mobile sensors with checkpointing constraints.** In *Proceedings of the 8th International Workshop on Sensor Networks and Systems for Pervasive Computing (PerSeNS, in conjunction with IEEE PerCom)*. Lugano, Switzerland, March 2012.

O. Saukh, D. Hasenfratz, C. Walser, and L. Thiele. **On rendezvous in mobile sensing networks.** In *Proceedings of the 5th Workshop on Real-World Wireless Sensor Networks (RealWSN)*. Como Lake, Italy, 2013.

O. Saukh, D. Hasenfratz, and L. Thiele. **Route selection for mobile sensor nodes on public transport networks.** In *Journal of Ambient Intelligence and Humanized Computing*, 5(3). Springer, 2014.

Curriculum Vitæ

Personal Data

Name David Hasenfratz
Date of Birth November 26, 1984
Citizenship Swiss, Hungarian

Education

2011–2015 ETH Zurich
Computer Engineering and Networks Laboratory
Ph.D. under the supervision of Prof. Dr. Lothar Thiele
2008–2010 ETH Zurich
M.Sc. ETH in Electrical Eng. and Information Technology
2004–2007 ETH Zurich
B.Sc. ETH in Electrical Eng. and Information Technology
2001–2004 Mathematisch-Naturwissenschaftliches Gymnasium
Bern-Neufeld
University-entrance degree (Matur)

Professional Experience

2011–2015 ETH Zurich
Computer Engineering and Networks Laboratory
Research and teaching assistant
2008–2010 Bain & Company Switzerland Inc., Zurich, Switzerland
IT-Support
2008 IBM Research Lab Zurich, Switzerland
Research intern
2007–2008 Virtusa, Hyderabad, India
Student intern

Awards

March 2014 Mark Weiser Best Paper Award at IEEE PerCom 2014