UiO : Department of Informatics
University of Oslo

Mobile Air Quality Monitoring for Exposure Estimations

Master's thesis

Eivind Engesæter

03-08-2015

Mobile Air Quality Monitoring for Exposure Estimations

Master's thesis

Eivind Engesæter

03-08-2015

Department of Informatics

University of Oslo

2015

Mobile Air Quality Monitoring

for Exposure Estimations

Eivind Engesæter

http://www.duo.uio.no/

Trykk: Reprosentralen, Universitetet i Oslo

Summary

Air pollution in the urban environment has become a factor of great impact on the human health. Both gaseous pollutants and airborne particle matter has been linked to a number of diseases, and are especially harmful to people with reduced health. World Health Organization[1] recently stated that air pollution is "*the world single biggest environmental risk*".

To gather knowledge on the actual levels of various air pollutants, monitoring stations strategically placed measure the level of pollution in the area. The monitoring data is also used in modeling to assess the levels of pollution in the areas not covered by the stations. Although this is very useful for general assessments, the spatial dimension is not covered. Thus, little is known about the distribution and variation of the air pollution in the urban environment.

Mobile monitoring seeks to add this knowledge, by employing mobile monitoring equipment to gather data on the air pollution in the urban environment. Low-cost sensors are now available to monitor air quality, and can provide high-resolution maps of the levels of air pollution.

This thesis use the prototype mobile monitoring platform developed by the Citi-Sense MOB project to monitor the levels of gaseous pollutants at Majorstuen in Oslo. By cycling predefined routes, the campaign seeks to gather knowledge on the spatial and temporal variations and distribution of air pollution in the area.

Due to issues with the sensors, the data from the campaign proved to be of so low quality that no knowledge could be obtained. Even though, the campaign provided insights that can be useful to other similar projects on mobile air quality monitoring.

Other studies have obtained high-resolution data, by using professional monitoring equipment. Such data, when available, can inform the public about the actual levels in their environment. This thesis proposes a system that uses high-resolution air pollution data and

[1] http://www.theguardian.com/environment/2014/mar/25/air-pollution-single-biggest-environmental-health-risk-who

user preferences to visualize the exposure and recommend healthier alternative travel routes for the users.

Preface

First I will thank my supervisor Josef Noll for the guidance in writing this thesis, and the visions and thought experiments we have had during the progress of the thesis. In addition, I will thank Nuria Castell of NILU for the supervising guidance and insights into the field of air quality monitoring that was all new to me before the thesis.

Secondly, I will thank the Citi Sense MOB project for providing the electrical bike with the monitoring equipment used in the monitoring campaign for the thesis.

Most of all I will thank my dear wife Katrine, my lovely daughter Lydia, my good dog Biff and my whole family for backing me up, providing support in hard times, and generally raising my life quality.

Picture 1 - Out monitoring

Table of Contents

List of figures, tables and pictures

Figures

Tables

Pictures

1 Introduction

This chapter introduce the domain of air pollution and monitoring. It will look into the history and the governmental actions taken to decrease emissions of air pollutants. Then it moves on to the concepts of smart cities, quantifiable self and community air pollution sensing projects. Last, the chapter explain the scientific methodology of this thesis and campaign.

1.1 Historic perspective

"Look back to look forward"

Industrial revolution

Ever since the start of the industrial revolution in the late 18th century, there has been an increasing use of hydrocarbon fuels, such as wood, coal, oil and gas. Since the dawn of age men has used wood as a source of heat or energy, but the earliest significant increase of fossil fuel use came with the introduction of the steam machine. The steam machines made it possible to operate various factories, but their large energy needs affected the environment in new ways, in the forms of industrial air and water pollution[2].

Along with the industrialization came the growth of big cities. New developments in agriculture lead to higher yield and less need for manual labor. People from rural and agricultural areas had to move to the cities in huge numbers to get jobs in the new factories. All these new residents needed to use wood as for their source of energy for cooking and heating. The huge need for wood lead to deforestation and shortages, so there was a need for a new alternative. Coal soon became the preferred source of energy of the industrial cities, both for the industry and the general population.

The combination of increasing levels of pollution from industry and private households made a new phenomenon appear, smog. The use of coal created sulfur-laden dense smoke that made the sky appear smoky and gray. People claimed that this smoke had harmful effects on

[2] http://www.eh-resources.org/timeline/timeline_industrial.html

clothes, corroded buildings, water quality and human health. This harmful smoke, combined with the fog, created a layer of gray polluted air over the city of London. This thick layer were given the name smog. The air pollution from smog caused rise in the death rates, and in a week in 1873, the smog in London killed more than 700 people. Yet, the worst air pollution disaster was to come. In December 1952 the Great London Smog killed about 4000 people, and an additional 100000 were made ill by the health effects of the smog[3] .There were recorded Sulphur dioxide and smoke concentrations of several thousand micrograms per 3^m (Brunekreef & Holgate 2002). Recent studies have shown that in December 1952 and January and February 1953, the mortality was as high as 13500 more deaths than normal, and the smog caused a large fraction of these deaths (Bell et al. 2003).

Picture 2 – Great Smog of London[4]

Rising concerns

The Great London Smog triggered great concern throughout the population and caused considerable pressure to take action against the problem. As a response, the Parliament of the United Kingdom passed the Clean Air Act in 1956. The Clean Air Act tried to reduce the air pollution by different measures, such as relocation of heavy industry and power stations away from densely populated areas, create areas permitting only smokeless fumes, and changing the domestic energy use from open-hearth coal fires towards cleaner sources[5].

[3] http://www.eh-resources.org/timeline/timeline_industrial.html
[4] http://blog.iso50.com/25702/the-great-smog-of-52/
[5] http://en.wikipedia.org/w/index.php?title=Clean_Air_Act_1956&oldid=637131238

In the US there were also raising concern about the rising air pollution and the environmental and physiological effects from it. In 1963, the Congress passed the US Clean Air Act, which gave funding to research air quality, and find measures to control the pollution. However, it was not until 1970 when Congress passed a new Act, that the federal government's role shifted towards a more regulative one, and initiated national programs to limit the emissions produced[6].

Since then there has been several new acts passed by governments around the world, and the smog problem has lessened in the western part of the world, but air pollution emissions is still prevalent in cities around the world. In some parts of Asia, especially in large cities or areas close to heavy industry, the air pollution levels are constantly at levels deemed hazardous by the WHO.

Today many homes install new clean burning ovens and heavy industry move away from cities. This has lessened their impact on the air quality, but wood burning is still one of the main sources of pollution in urban areas. The tremendous growth of cars and other vehicles powered by hydrocarbon fuels in the last 50 years has made traffic one of the main contributors to the urban air pollution. In addition to the air pollution emitting from the combustion, the motor vehicles also produce harmful particle matter, especially from spiked tires ripping up the tarmac.

The visible air pollution, such as black smoke, has decreased significantly since the 1960s and 70s due of the measures taken to reduce the emissions. Although this is positive for the population, the levels of invisible pollution has risen. For instance, one can find high levels of ozone concentration not only in warm and sunny places as Los Angeles and Mexico City, but also in large parts of Western Europe. Levels of nitrogen oxides has also increased, mainly due to the ever-rising number of motorized vehicles. In addition, the toxicity of airborne particles has altered due to changes in their size distribution and composition (Brunekreef & Holgate 2002).

[6] http://www.epa.gov/air/caa/amendments.html

Harmful traffic emissions

There has been various efforts made to reduce the harmful effects of the traffic. Petrol no longer contain lead. New motor standards as Euro6 makes for more effective and fuel-efficient engines, and thus produces less pollution. Electric and hybrid vehicles are starting to become popular, especially in Norway where incentives given by the government has increased the demand for these types of motor solutions.

To limit the emissions of the traffic, the EU has created standard maximum emission levels for all new vehicles. Since 2009, all new light traffic such as cars and light transport must adhere to the Euro5 standard. This standard has recently gotten even stricter with the Euro6 standard, which further decreases the accepted emissions of NOx for diesel engines. These classifications also applies to heavy transport, but the accepted limits are higher, and is instigated a year later than the light traffic[7].

1.2 Public action

Due to the raised concerns and awareness on the effects of air pollution, various research institutions has been founded around the world. In Norway, Norsk Institutt for Luftforskning[8] (NILU) was established in 1969, as an independent institution for air quality research. Their aim is to increase the understanding of the processes and effects associated with the composition of the atmosphere, climate change, air quality and environmental toxins.

The Organization for Economic Cooperation and Development[9] (OECD) has recently published data showing that the levels of air pollution in western Europe has fallen for some pollutants, but still is well above the recommended levels.

Global projections

OECD has also made projections for estimating the concentration levels in 2050 if no new policies are introduced to reduce the risks[10]. The urban air quality will continue to

---

[7] https://www.regjeringen.no/nb/sub/eos-notatbasen/notatene/2007/feb/euro-5-og-6-avgasskrav-for-lette-kjoreto/id523447/
[8] http://www.nilu.no/OmNILU/tabid/67/language/nb-NO/Default.aspx
[9] http://www.oecd.org/env/the-cost-of-air-pollution-9789264210448-en.htm

deteriorate on a global level and outdoor air pollution such as particulate matter and ground-level ozone is expected to become the top cause for environmentally related deaths.

The concentration of air pollution has in some cities, especially in industrial regions of Asia, has risen far above the levels satisfactory by the WHO Air Quality Guideline. As this situation is likely to continue, governments must undertake significant efforts to reduce the health impact.

In key emerging economies, the emissions of sulphur dioxide (SO_2) and nitrogen dioxide (NO_2) will most likely increase substantially. The projections are 90% higher for SO_2 and 50% higher for NO_2 compared to year 2000. In addition, the deaths related to particulate matter are likely to be more than doubled, especially in China and India. Premature deaths due to ground-level ozone will also likely be highest in China and India.

The global projection towards 2050 is that the increasingly aging population, along with the effects of the urbanization, will outstrip the possible benefits of any emission reductions. The current baseline levels from the most polluted areas are currently so high that a predicted 25% decrease in emissions will have low impact on the number of deaths related to air pollution.

In the OECD countries however, having a long-standing commitment to improve air quality along with the means to enforce it, has had a reduction in the emissions of SO2, NOx and black carbon over the last decades. OECD expects this trend to continue towards 2050. However, the levels of particulate matter (PM) and ozone (O_3) is still higher than the accepted thresholds. Despite efforts made to reduce the amounts of precursor gases of ozone, the concentration in Europe are relatively stable (EEA 2014). The variety of factors to account for, and the air-borne nature of the pollutants, increase the challenge to find viable methods to reduce the exposure.

Acceptable levels set by authorities

The local air pollution levels in Norway has recently gotten both public and political attention. Research has found that in the days after a rise in the levels there is elevated

[10] http://www.oecd.org/environment/indicators-modelling-outlooks/49928853.pdf

number of deaths due to cardio-vascular diseases[11]. The media has brought this to attention for the public, and there has been extensive coverage of dangerous levels of air pollution in Oslo this winter (2014/2015) [12] [13] [14].

An investigative report done by NILU and Transportøkonomisk Institutt (TØI) (Høiskar et al. 2014) for the City Council of Oslo states that the measured levels of coarse particle matter and nitrogen dioxide exceeds the acceptable thresholds a number of days each year. It also evaluate new reduced thresholds on the current levels, and propose a set of actions to reduce the level of air pollution.

The EEA (European Economic Area), which Norway is a member of due to the EFTA agreement, has levels for acceptable pollution set by their legislation. In 2014, The EFTA (European Free Trade Association) sent a complaint to the Norwegian authority addressing the fact that the levels measured in various cities and parts of Norway exceeded the levels set by legislation. If the Oslo City Council do not take sufficient actions to reduce emission levels, ESA, the legal branch of EFTA, threatens to take the municipality of Oslo to court because of the number of breaches of the thresholds.

The City Council of Oslo has proposed, as a mean to reduce the levels of urban air pollution, to prohibit driving diesel cars in parts of or the whole of Oslo. Based on the levels obtained by fixed-site monitoring stations the prohibition will take effect when the levels are above the acceptable thresholds for two days or more. Furthermore, they propose lower speed limits, and differentiated road pricing to be added the measures for reducing the emissions.

1.3 Smarter cities

Cities around the world are now utilizing the combination of sensors, communication and data processing to apply active real-time assessment and management of several aspects of the city infrastructure. This can be seen as knowledge infrastructure, and spans widely into areas such as intelligent transport systems, garbage and sewage handling, water

[11] http://www.vg.no/nyheter/innenriks/bil-og-miljoe/ny-rapport-doedelig-daarlig-oslo-luft/a/23357361/
[12] http://www.vg.no/nyheter/innenriks/bil-og-miljoe/farlig-luftforurensing-i-oslo-her-lyste-det-roedt/a/23357385/
[13] http://www.nrk.no/ostlandssendingen/livsfarlig-luft-i-oslo_-_-jeg-kjenner-et-press-mot-hjertet-1.12107301
[14] http://www.osloby.no/nyheter/Med-kulden-kommer-helsefarlig-svevestov-7433170.html

consumption, law enforcement, healthcare and air quality monitoring. The incentives for the cities is to enable better services and information for the citizens, while also reducing the cost of providing the services. This knowledge infrastructure do not only include the equipment and communication infrastructure components, but also the services that handles and analyzes the information, and finally visualizing it to the users.

As this knowledge infrastructure is becoming a reality, it is enabled by several information technology advancements. Recent year's development in the fields of sensors and Internet of Things, big data and analytics and mobile networks has improved the technologies to sense and handle the immerse stream of data provided by this infrastructure.

The handling of these amounts of data through cloud services is one of the most strategically important areas for several of the world's largest computing businesses. The price for cloud computing resources have plummeted, and the services provided also include the availability of advanced analytics and prediction and recommendation systems. Examples of this are IBM's supercomputer system Watson[15] and Amazon's AWS Machine Learning platform[16], both available as cloud-based services. The lowered cost and service availability of the cloud platforms will reduce the need for building and maintaining this part of the infrastructure, and speed up implementations.

The consulting company Navigant Research[17] envisions the market of smart city technologies to exceed 174 Billion USD from 2014 to 2023, with the Asia pacific as the largest market due to rapid urbanization. Pike Research[18] forecasts similar numbers, expecting an annual growth of 16 Billion USD.

1.4 Quantifiable self

Meanwhile the cities are getting smarter; the citizens have started sensing their environment and context themselves. Building on the advancements in technology mentioned earlier, along with the capabilities of modern smart phones, the citizens are

---

15 http://www.ibm.com/smarterplanet/us/en/ibmwatson/what-is-watson.html
16 http://aws.amazon.com/machine-learning/
17 http://www.navigantresearch.com/newsroom/investment-in-smart-city-technologies-is-expected-to-exceed-174-billion-from-2014-to-2023
18 http://www.greenbiz.com/blog/2011/09/29/market-smart-city-technology-reach-16b-year-2020

utilizing a wide array of sensors in body worn technologies to provide quantifiable data. Activity trackers, depending on the functionality of equipment, are able to measure activity, sleep patterns, heart rate and other quantifiable variables. The data is analyzed against user-set goals, and visualized to the users on their mobile apps. This functionality is also available in the latest versions of smart watches. Apple Watch[19] is now available to several markets, and one of the core functionalities is health monitoring.

Most of the mobile apps in this domain has the possibility to share the data and statistics trough social media platforms. While one should be aware of the dangers of sharing personal data to such networks, it can also encourage use and promote a healthier lifestyle in their community.

Smart clothes are clothing designed to sense the environment and alert the user if there is changes that the user should be aware. Extreme environments such as the Arctic calls for equipment that is able to alert the user before the conditions become dangerous. Much of the same sensors and techniques found in the equipment already mentioned is used in this new context. Sintef is now working on a project named SmartPro[20] to create intelligent clothes for these extreme environments.

1.5 Recommender systems

The information available to the public regarding air pollution is not always directly understandable, as most people is not aware of the health impact of the various pollutants and the measured level of concentrations. Recommender systems has traditionally been developed to suggest personalized alternatives to the user on items or content, and to alleviate the problem of information overload. Recent years the recommender systems has been given the opportunity of adding contextual information to the recommendations, to further improve the personalization.

Mobile monitoring campaigns has the potential to inform the public of the actual levels of harmful pollutants in their environment. While this information might be of high interest,

[19] https://www.apple.com/watch/
[20] http://www.sintef.no/nyheter-fra-gemini.no/utvikler-intelligente-ekstremklar/

the information on levels of pollutants has to be easily understandable. Visualization of pollution levels on maps will give a general description, but a more personalized approach tailored to the user preferences will give an immediate notion of value to the users. A recommender system for the air pollution data can incorporate the publicly available data with personalized preferences to tailor the recommendations for the users.

An approach for a context-aware recommender system for selecting a route less polluted than the fastest route will be presented in this thesis, building on data obtained from a mobile monitoring campaign situated in Oslo.

1.6 Community sensing projects

Recent advancements in ICT technology has opened up a new field of air quality monitoring. The availability of low-cost sensing technologies enables production of air quality monitoring equipment with a significant lower price than the traditional equipment used. Additionally, as the reduced size and power consumption of the equipment enables the utilization in mobile monitoring, it is able to produce air quality data with unprecedented temporal and spatial resolution. Several ongoing research projects are developing novel sensing equipment along with tools to analyze and visualize the data.

1.6.1 The EveryAware project[21]

This project, consisting of a consortium of researchers from complex systems, environmental monitoring and modeling, and social sciences and computer science, seeks to enhance the public environmental awareness through social information technologies. Their aim for the project is to provide capabilities for environmental monitoring, data aggregation, and information presentation to the users by means of mobile and web-based tools such as smart phones, computers and sensors.(EveryAware White Paper n.d.) By leveraging on the low-cost and high usability of the sensor platform developed for the project, they seek to enhance the awareness and create behavioral change from the personalized environmental monitoring.

[21] http://www.everyaware.eu/

1.6.2 The OpenSense project[22]

The OpenSense project has similar aims as the EveryAware project, but focus on understanding the health impact of exposure to air pollutants, and provide high-resolution air quality maps, along with the potential of using crowdsourcing to provide feedback to its users. By adding mobile and crowd-sensed data sources to the already available air pollution data they seek to increase the spatial and temporal resolution of the air quality maps.

1.6.3 The Citi-Sense MOB project[23]

This project seeks to raise the awareness of urban air pollution, its impact on human health and the benefits of using transportation with less exposure to air pollution. By gathering data from mobile monitoring the project aim is to continuously monitor, communicate and share the environmental data, and to contribute to the city infrastructure and real-time city management and sustainable progress.

The project relies heavily on involvement from the urban community, and seeks to motivate the public to contribute in the monitoring campaigns. People will be equipped with low-cost monitoring equipment and smart phone applications to monitor and assess their personal exposure, and to share their experiences with the community. Gamification and Fun theory concepts will be applied to help motivate the public and possibly join the project.

This thesis will contribute to this project by doing a mobile monitoring campaign in the urban environment with the bicycle monitoring platform developed by the project. The campaign seeks to investigate how the levels of pollution in the urban environment varies depending on the time and location of monitoring, and how the levels change while moving in this environment. It will also look into the current state of the monitoring hardware, to assess its functionality for such monitoring campaigns.

[22] http://opensense.epfl.ch/wiki/index.php/OpenSense_2
[23] http://www.citi-sense-mob.eu/Home.aspx

1.7 Scientific methodology

This thesis and monitoring campaign will follow the scientific methodology proposed by Glass (1995 p. 3), stating the methodology as "*Observe the world, propose a model or theory of behavior, measure and analyze, validate hypotheses or the model or theory, and if possible, repeat"*

The methodology has the following four phases:

- **Informational phase:** gathering and aggregating information via reflection, literature survey, people/organizational survey or poll.

- **Propositional phase:** proposing and/or formulating a hypothesis, method or algorithm, model, theory or solution.

- **Analytical phase:** analyzing and exploring a proposition, leading to a demonstration and/or formulation of a principle or theory.

- **Evaluative phase:** Evaluating a proposition or analytic finding by means of experimentation (controlled) or observation (uncontrolled, such as a case study or protocol analysis), perhaps leading to a substantiated model, principle or theory.

This thesis use the phases in the following way:

Informational phase

The informational phase consists of a literature review of mobile monitoring campaigns that has used bicycles as the transport vehicle. The review focus on the methods employed to conduct the various campaigns, such as route selection and monitoring equipment. It will also review the conclusions and issues faced in the campaigns. Finally, it will present a summary and methods for the campaign.

Propositional phase

The propositional phase consists of a description of the campaign questions and challenges.

Analytical phase

This is the largest phase, containing both the methods for the campaign and the actual campaign execution, as well as the analysis of the data from the campaign.

Evaluative phase

The last phase will address the campaign questions and challenges in the light of the data and experiences provided by the campaign and the literature review.

2 Recommender systems

This chapter will introduce recommender system models in general, and the different categorizations. Then it explores possible methods of adding contextual information to make the recommender system context-aware. Last, it will examine challenges encountered with such systems.

2.1 Traditional recommender systems

A recommender system is a software application that is able to propose personalized recommendations to the user based on their current preferences and their earlier choices. Nicolas Negroponte envisioned and proposed the first personal recommender systems in the 1970s, and in the early 1990s Goldberg et al. conceptualized and prototyped the first recommender system named Tapestry, and in the mid-1990s implementation and commercialization were presented by Resnick and Varian (1997).

Since then different recommender systems has become an integral part of the services provided online, in a multitude of domains. The largest e-commerce sites are now providing its users with recommendations on their offered services and items. The recommendation systems is implemented for their ability to offer more services or items to the user, driving up both user satisfaction and increased revenue for the provider. In addition to this, the systems try to negate the challenges with the information overload often found on e-commerce sites.

2.1.1 Information overload

Information overload is a phenomenon where the available sources of information becomes a hurdle instead of helpful for the user, as the user struggles to find the relevant content or items for their interest. This might cause the user to stop and go to another vendor that enables the user to locate relevant content or items. A good recommender system along with appropriate information architecture will increase the user satisfaction and strengthen the relationship between customer and vendor (Schafer et al. 1999).

2.1.2 Personalization and user profiles

A common goal for recommender systems is to provide personalized information to the user. Recommender systems use user profiling extensively to gain knowledge about the user, and to provide more effective, faster and satisfied goal achievement (Nadolski 2009). This enables the recommender system to select and prioritize the most relevant information adaptively to the user profile. As the recommender system can be seen as a tool for generating recommendations by creating and exploiting user profile, the user profile will always have a central role in these systems (Tintarev & Masthoff 2011).

There are various methods used in recommender systems to generate user profiles. Direct user input is a common starting point where the user explicitly gives personal information in various categories. Data mining techniques are able to attain data on user behavior history and patterns, and to select relevant information based on this historical data. Moreover, the user profile might include the different relations between users to improve the relevance and accuracy of the recommendations.

2.1.3 Recommender system models

Traditional recommender systems has historically been categorized into different models depending on the aim of the recommendations and the data available. These are content-based models, collaborative models, knowledge-based models, demographic models, community-based models and hybrid models.

Hybrid models is various combinations of the earlier mentioned models. Depending on the goals of the system, they utilize models in varying degree to be able to exploit the strengths of both models.

Content-based models

The content-based models take a Top-down approach, which is to model the domain knowledge for the system by relying on expert knowledge and involvement, either in a static or dynamic way (Manouselis et al. 2011). These systems often try to recommend items or content that is similar to earlier user choices.

The content-based models often use a three layered architecture model (Manouselis et al. 2011). This model has a knowledge representation layer, an adaptation layer, and an interface layer. The knowledge representation layer splits into three resource pools, which is utilized by the adaptation layer to make the recommendations.

These three pools contain information on the domain knowledge, the user profile, and relevant data. The user profile defines the information on the user specifics, such as personal data, their preferences and their interests. The domain knowledge describes the specific domain for the system, often defined in an ontology or a taxonomy of terms. Information resources hold general relevant information so the system is able to make recommendations.

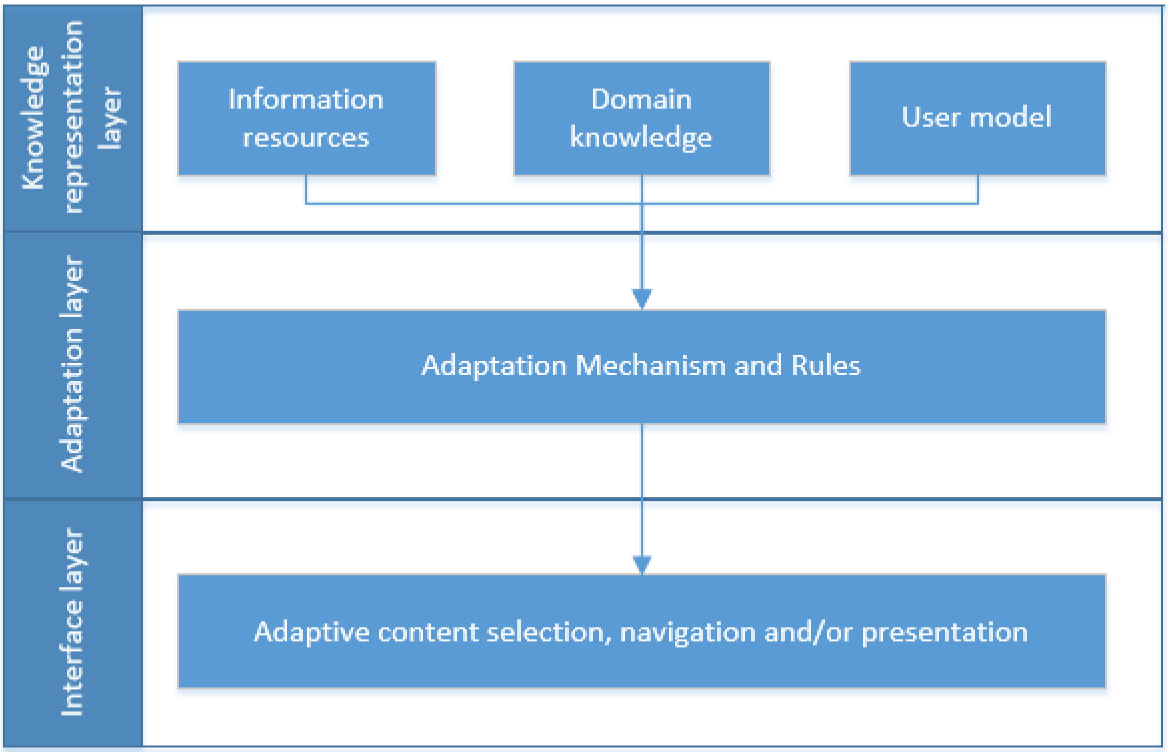


Figure 1 - Content based three-layered architecture

These three knowledge pools are processed by the adaptation mechanism and rules to meet the goals of the recommender system of proposing the best personalized recommendations. These recommendations are processed again in the third layer. This layer deals with data presentation, and displays content adaptively to the interface of the user.

Collaborative models

Collaborative models take the opposite approach, by letting the system employ a bottom-up approach. By using techniques from the field of artificial intelligence, such as machine learning and data mining, the system are able to develop filtering mechanisms to make the domain model and to provide recommendations.

The collaborative models are relying on the participation of its users to gather knowledge. It uses the concept of emergence to let the individual actors and resources self-organize to shape higher-level patterns of behavior (Manouselis et al. 2011). These patterns of behavior can then be conveyed to other users of the system to improve the relevancy of the recommendations. These models do not rely on domain modeling and knowledge, as they grow from bottom-up and form the domain knowledge from the information provided.

Through the rise of social media collaborative models and filtering has gained a strong foothold, and is used to a large degree for proposing content and making recommendations in these networks.

Knowledge-based models

The knowledge-based models rely on specific domain knowledge on how certain items or services are able to meet the needs and preferences of its users (Tintarev & Masthoff 2011). These systems can be sub-divided into case-based and constraint-based systems, where their main difference is how the recommendations is calculated. Case-based systems estimates how the solution of the problem match the description of the problem, e.g. how the user needs match the recommendations given. Constraint-based systems utilize predefined knowledge bases containing rules on how user requirements relate to item features.

Demographic models

This model is based on the assumption that recommendations should be created to match different demographic niches. For example, the user's language or country could be used to improve the personalization of the recommendations.

Community-based models

The community-based models creates recommendations based on the preferences of the friends of the user. The rise of social media where friends connect and share their preferences enables simple but still comprehensive gathering of data created from the social relations and preferences of the users.

Hybrid models

All of the aforementioned models has their strengths and weaknesses, and to harness the strengths and minimize the weaknesses, hybrid models employ elements of different models. An approach named "collaboration via content" are based on the techniques of the collaborative models, but additionally maintain content-based profiles of the users (Adomavicius & Tuzhilin 2005).

Empirical studies has compared the performance of hybrid models to the content-based and collaborative models, and have demonstrated that hybrid methods are able to provide more accurate recommendations than only selecting one model (Adomavicius & Tuzhilin 2005).

2.2 Context-aware recommender systems

Traditional recommender systems available to the public has been inclined towards achieving business goals and short-term sales (Werthner et al. 2007). As their objective often is to increase the sales of items or use of a service, they can in fact increase the problem with information overload.

Contextual information might be able to mitigate the problem by adding this information to the recommender system. This can be information on the location and time obtained implicitly and automatically from the user, but also information stated explicitly by the user, such as mood and interest limitations. Additionally, behavioral research in marketing has found that the decision making of users are contingent on the context. (Adomavicius & Tuzhilin 2011)

The availability of contextual information has had a huge increase as most of the mobile phones today offer location data through GPS coordinates. These coordinates provides spatial information, and time data provides temporal data. Additionally, the phones often enable access to user information such as calendars, emails and notifications. This contextual information can be added to recommender system to increase the quality and relevance and make contextually appropriate recommendations (Werthner et al. 2007).

Werthner et al. (2007) calls this context and smart phone enabled recommender systems the third wave of recommender systems. On-demand cloud computing resources, recommendation engines as services and online data sources offered through APIs are all enablers that lower the cost and make it easier for developers to create recommender systems on massive amounts of data. This also enable the developers to have more focus on the user experience and less on the technical side of algorithms.

The emphasis on user experience and Human Computer Interaction (HCI) in recommender systems has grown the recent years, and researchers are arguing that in addition to the accuracy and effectiveness of the algorithms, the user acceptance of the systems and its recommendations should also be evaluated (Tintarev & Masthoff 2011). This has gained additional interest, as the usage growth of mobile interfaces has been tremendous recent years, and context-aware recommender system benefit greatly of the inherent features of mobile platforms.

2.2.1 Context

The concept of context is multifaceted and many definitions exists, depending on research paradigm and discipline. Schilit & Theimer (1994) define context as "*location and the identity of the nearby people and objects*. A broader perspective is presented by Tintarev & Masthoff (2011 p.7) "*Context is defined with a predefined set of observable attributes*".

The notion of context in recommender systems has two major characteristics and inherent advantages, being improved personalization and ubiquitous computing (Zhang & Wang 2010). Ubiquitous computing refers to an "*any time, any place, any way*" access to information and computing resources, and improved personalization is to refine the recommendations with contextual information.

Context-aware computing is also known in the research community as pervasive computing, ubiquitous computing, embodied interaction and more (Dourish 2004). The hope of context awareness is to make the systems more aware of the specific settings of use.

For context aware recommender systems, the context information can be obtained from a variety of sources, such as mobile phones or wearables such as fitness monitors. A modern mobile phone has several sensors to provide contextual information, such as gyroscope, accelerometer, proximity sensor, ambient light sensor, moisture sensor and compass. These sensors provide additional information on the context of the user, extending the contextual information on location and time. There will probably be a plethora of devices able to provide such contextual information with the coming of the Internet of Things.

2.2.2 Obtaining contextual information

To gather the contextual information from the user there a set of techniques employed (Verbert et al. 2012; Adomavicius & Tuzhilin 2011).

- **Explicitly**
  Data collection take place through directly questioning the users, or eliciting the information through other means. Examples are registrations or other mandatory web forms requiring user input.

- **Implicitly**
  Implicit data collection take place when the environment changes, for example when the system detect that the user has changed location, or the timestamp of a transaction has changed. There is no need for interaction from the user as the system gathers the contextual information directly, and extracts the desired information.

- **Inferring**
  Inferring is applying methods from the fields of statistics or data mining to gather information not directly available, but inferred through matching user patterns to user profiles. To be able to infer contextual information, one must build a predicative model and train it with appropriate data.

2.2.3 Algorithmic paradigms

To introduce the contextual information to the recommendation process, Adomavicius & Tuzhilin (2011) propose three possible paradigms. Based on the same sources of data, the recommendation systems incorporates the contextual information at different stages of the recommendation algorithm.

Contextual pre-filtering

The addition of contextual information prior to the recommendation process provide the ability to pre-filter the recommendations, and select or construct the most relevant dataset for the given context. All of the aforementioned recommendation techniques can benefit from pre-filtering on contextual information.

Context generalization

To cope with the “sparsity problem” often encountered in recommender system literature, that being that the systems do not have enough data points to make relevant recommendations, Adomavicius et al. (2005) introduce the idea of generalized pre-filtering with large contextual segments. A superset of the given context is denoted into a contextual segment, based on the context taxonomy and the context granularity desired. Through a manual expert knowledge approach, or automatic empirical-based discovery, the best segment filter for the given context is applied to the pre-filtering process.

Contextual Post-filtering

For this paradigm, the recommendation is performed at the entire dataset, without any contextual information. Then, after the general recommendations are ready, they are filtered to the context of the users. As the filter process is performed after the general recommendations is done, it is also able to handle the different recommendation techniques. The basic notion of post-filtering is to analyze contextual preference information for a specific user in a specific context to discover specific usage patterns applicable to the filtering process.

The filter process can be divided into two techniques:

Heuristic – this technique focus on common item attributes for a given user in a given context. These attributes are used to filter out items lacking the desired attributes, or to rank the recommendations based on the attributes.

Model based – the model based technique approach is to build predictive models to calculate the probability of relevance to the user selection of a certain item at a given context, and adjust the recommendations based on the relevance. A filter approach is taken on items that have a probability of relevance below a given minimal threshold, or a ranking is done by weighting the probability of relevance against the predicted rating.

Contextual Modeling

For this paradigm, the contextual information is added directly into the recommendation model, and is thus an integral part of the recommendation system. This paradigm make the recommendation process multidimensional, as the context is added along the dimensions of users and items. Predictive models built using techniques such as decision trees, regression or probabilistic models, or heuristic calculations incorporating contextual information to the user and item data are used for the contextual modeling.

On the next page is a graphic describing the algorithms, and at which stage the recommendation algorithm integrate the contextual information

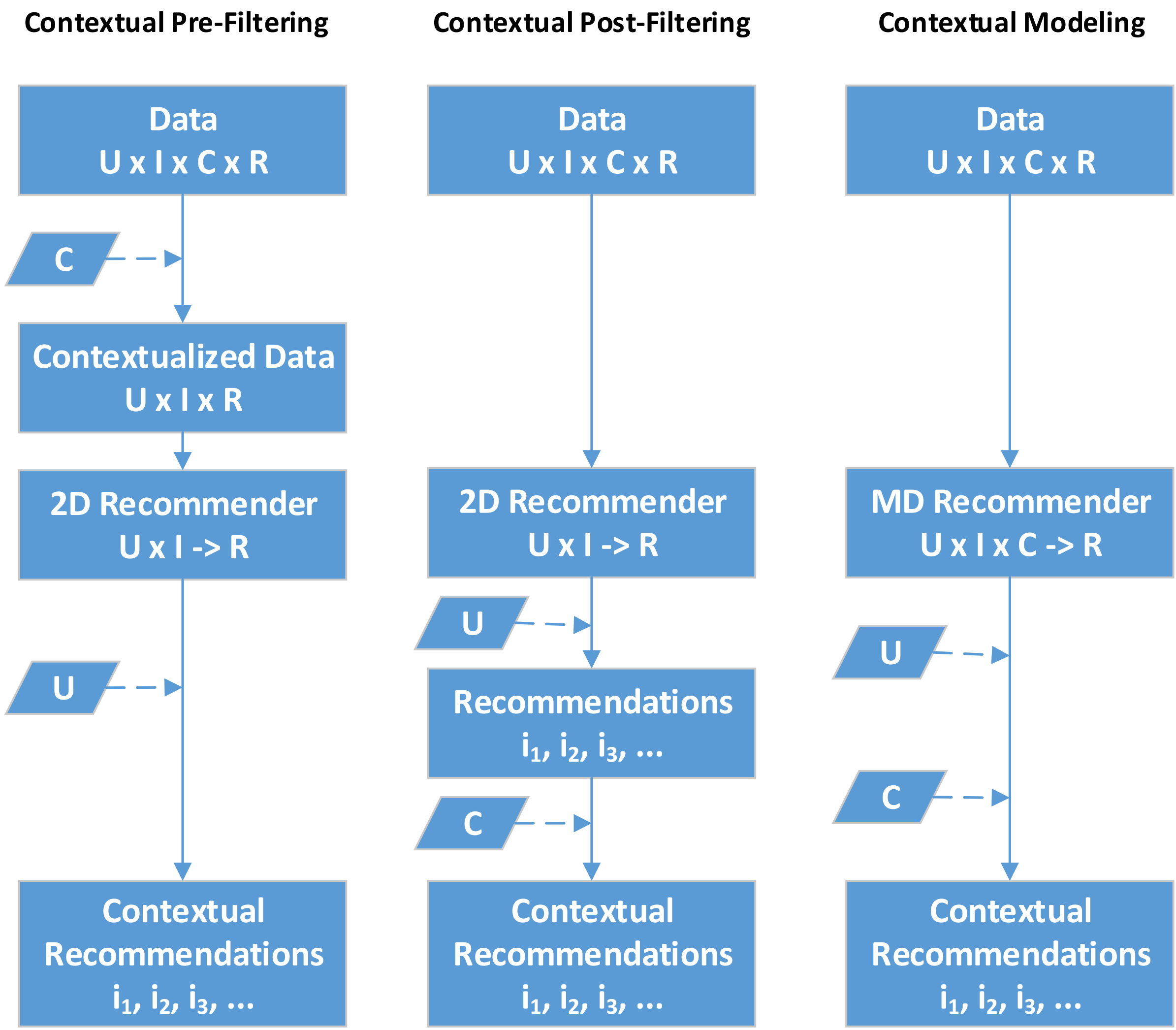


Figure 2 - Overview of the algorithmic paradigms – Adopted from Adomavicius & Tuzhilin (2011) . User = U, Item = I, Context = C, Rating = R, Recommendation = i_n*.*

Combining the approaches

Combining several context-aware recommendation techniques into a blend or ensemble of the different solutions can often provide significant improvements over individual approaches. This blending can be achieved by developing and combining several models of the same sort, for example a set of different contextual pre-filters to handle various generalizations of the same context.

Contextual information might also be divided into separate components, and the utilization of the components contextual information might differentiate depending on the use in pre-filtering, post-filtering or modeling approaches.

Although the selection of algorithms is an important factor in recommender systems, Werthner et al. (2007 p. 23) states that *"algorithmic precision is just one of many factors that affect a user's adoption of a recommendation, and other issues such as interface design, long-term performance evaluation, or context-awareness are also prominent parts of a recommender".*

2.2.4 Challenges

Recommender systems has a number of challenges important to the development and research of such systems.

Combination of models

Combining different models has been successful in many project, but the performance and scalability of the different combinations still need further research (Adomavicius et al. 2005).

Lack of publicly available datasets

A main challenge in development of recommender systems, and specifically for context-aware, is the general lack of available datasets to the public. This make conducting experimental evaluation of context-aware recommender systems a challenging task (Zhang & Wang 2010).

Sparsity problem

A common problem for some of the recommender system models is that they have very little information available compared to the data needed for providing predictions (Zhang & Wang 2010). Hybrid models employing collaborative filtering can possibly overcome this problem by a collaborative approach (Adomavicius & Tuzhilin 2005), or to make models more robust and scalable in making more accurate estimations where rating data is scarce (Adomavicius & Tuzhilin 2011).

Filter bubble

A problem of employing recommender systems is the possibility of the users getting caught in a "Filter bubble". This term, coined by Eli Pariser, describe the situation where *"online personalization effectively isolate people from a diversity of viewpoints or content"* (Nguyen et al. 2014 p. 677). This can lead to narrowing the recommendations to a level that do not account for the actual preferences of the user, and development of recommender systems should be aware of this preference narrowing effect.

3 Domain Knowledge - Urban Air Quality

This chapter seeks to gather domain knowledge in the field of urban air quality. This field spans several research domains, such as medicine, meteorology and environmental research. First, there are reviews of various harmful pollutants found in the urban environment, and the implications of pollution to human health. Secondly, the thesis examines various factors influencing the variations in urban air pollution.

The field of urban air quality and health effects is highly complex, with many contributing external factors, and contradicting research results. Therefore, this thesis will only provide a shallow review, and a deeper review is necessary to do a full assessment of the domain knowledge and the relationships between the factors.

3.1 Air Pollutants

An air pollutant is by definition any substance that may harm humans, animals, vegetation or material (Kampa & Castanas 2008). Its effect on humans might be a present or potential hazard for human health, and cause or contribute to serious illness or an increase in mortality.

Air pollutants have several different characteristics, as their chemical composition, reaction properties, emission, time of disintegration, ability to transport for short or long distances and their impact on human health. Still, there is some similarities they share, which enables us to divide them into four groups:

- **Gaseous pollutants**
- **Particulate Matter**
- **Persistent organic pollutants**
- **Heavy metals**

3.1.1 Gaseous pollutants

The gaseous pollutants is a large group of gases comprising mainly from sources of fossil fuel combustion, that contribute to the composition variations of the atmosphere to great extent. These gases can under certain environmental conditions react with each other, and thus change the atmospheric content.

These gases include:

- **NO** – Nitrogen monoxide
- **NO_2** – Nitrogen dioxide
- **O_3** - Ozone
- **SO_2** – Sulphur dioxide
- **CO** – Carbon monoxide
- **VOC** – Volatile organic compounds

Nitrogen oxides -NO_x

The Nitrogen oxides is a sub-group of gases consisting of oxygen and nitrogen. The term NO_x refers to the gases nitrogen monoxide (NO) and nitrogen dioxide (NO_2). The NO concentrations found in cities is not harmful by itself, but reactions with Ozone will create additional NO_2. The levels of NO_2 measured in the larger cities in Norway has exceeded the levels set by the government, but it is seldom an issue in less populated areas.

Road traffic, especially from diesel combustion, together with industrial emissions and maritime traffic are the main sources of NO_x emissions.

Ozone (O_3)

Ozone, a color and odorless gas, is both beneficial and harmful. While the ozone residing in the stratosphere protects us from the UV radiation from space, ground level ozone is reactive and potentially hazardous to the local environment and public health. Ground level production of ozone is a result of chemical reactions between NOx, CO, VOCs and sunlight.

Carbon monoxide (CO)

Carbon monoxide is a colorless, odorless and tasteless gas. It is a byproduct of combustion or other burning processes where the oxygen supply is too low to produce carbon dioxide (CO_2). There are also emissions of CO from more natural causes, such as wildfires and volcanic activity. The level measured in populated areas are too low to have an impact on the public health.

Sulphur dioxide (SO_2)

Sulphur dioxides is formed through combustion processes involving Sulphur, mainly from oil and coal, together with various industrial processes. The main sources of emissions are power plants, oil refineries and large industries. In western countries SO2 emissions is mainly a local concern around the areas of large emission sources.

Both SO2 and NOx has the ability to transport in the atmosphere for long ranges. It is later reintroduced in other areas through acid rain and particles, and can be severely harmful for the environment.

Volatile organic compounds (VOC)

Volatile organic compounds (VOCs) are gases emitted from a variety of solids and liquids. They include a range of chemicals that might have unhealthy short- and long-term effects. Concentrations of VOCs can be up to 10 times higher indoors than outdoors, and is found in a variety is household products such as paint, cleaning agents, printers and many more. (EPA 2012). They are also a component of fuel and other oil based substances. By using these products, one can expose oneself to very high pollutant levels, and these high concentrations can persist in the air for a long time.

3.1.2 Particulate matter

PM pollution is a mix of solid and liquid particles suspended in the air, with variances in their size, shape, chemical composition, surface area, solubility and origin (Pope & Dockery 2006). The size and surface of the particles, their number and their composition are important parameters for determining their health effect impact (Kampa & Castanas 2008). The size of the particles varies from a few nanometers to tens of microns, and because of their size, they penetrate the respiratory system to varying degree. Generally, they are divided into three groups, namely PM10, PM2.5 and ultrafine particles.

The composition of the particles varies, since they can absorb and transfer a wide variety of pollutants. The majority of components present are metals, organic compounds, biologic materials, ions, reactive gases and the particle carbon core (Kampa & Castanas 2008).

Several Norwegian cities has levels of PM that is higher than the recommended levels set by the government[24]. There is uncertainty whether the thresholds are set at levels that are tolerable for the public. Studies have shown that estimated concentration-response functions are nearly linear, thus giving no evidence of safe threshold levels (Pope & Dockery 2006).

PM10

The PM10 particles has a diameter below 10 microns, and include the PM2.5 and ultrafine particles. They are also called coarse particles. In Europe, the PM2.5 constitutes 50-70% of the measured PM10 (World Health Organization 2013). They derive from sources such as abraded soils, road traffic, construction debris, wood burning, industrial processes or coagulation of smaller particles (Bernstein et al. 2004). Their size enables them to penetrate the lower respiratory system, and therefore called thoracic particles (Brunekreef & Holgate 2002).

[24] http://www.luftkvalitet.info/Theme.aspx?ThemeID=6fc2e3cd-424f-4c03-ad0c-2b9c15369cd9

PM2.5

PM2.5 particles are smaller than 2.5 microns in diameter. The smaller size makes them able to penetrate further into the respiratory system, in the gas-exchange regions of the lungs. This gives them the description of being respirable particles (Brunekreef & Holgate 2002). The main sources of PM2.5 are wood burning, road traffic and industrial processes and long-range transboundary air pollution (Bernstein et al. 2004).

Ultrafine particles (UFP)

The ultrafine particles are particles smaller than 100 nanometer. Because of their size, they have low contribution to particle mass, but they are the most abundant in terms of numbers.

Urban and industrial environments are constantly receiving emissions of ultrafine particles as a result of combustion processes such as vehicle exhaust and photochemical reactions in the atmosphere (Thai et al. 2008; Pope & Dockery 2006).

Primary ultrafine particles have a very short life, but has the ability to rapidly grow into PM2.5 through coagulation and/or condensation (Thai et al. 2008)

Their small size gives them the ability to penetrate deep into the lungs, and may be more likely to be able to translocate from the lungs to the blood and thus reaching other parts of the body (Pope & Dockery 2006).

Black Carbon (BC)

Black carbon, popularly named soot, is a constituent of PM2.5 and consists of linked forms of pure carbon. Black carbon is often produced and emitted as an effect of incomplete combustion processes, and are thus closely linked to vehicle traffic (Peters et al. 2014).

Persistent organic pollutants

This group of pollutants (POPs) can persist in the environment for extended periods, having long half-lives in soils, sediments, biota and air. POPs can have a half-life of years ranging from years to decades in soil or sediments, and several days in the atmosphere (Jones & de

Voogt 1999). In addition, POPs are lipophilic chemicals that gets stored in the fat tissue, and as they move up the food chain, their effects accumulate.

The POPs can under appropriate environmental conditions enter a gas phase through volatilization from soils, vegetation and water bodies and enter the atmosphere. Then, due to the persistence, it can travel for long distances before it re-deposits.

Commercial products have added POPs to improve their effects in areas such as crop production, disease control and industry. However, these chemicals have had severe effects on the environment and on human health. Pesticides, dioxins, furans and PCBs belong to this group consisting of many thousand chemicals. Through tough regulations there has been great reductions in the use of POPs (EPA 2015).

Heavy metals

Heavy metal has gained ground as a general term for metals or metalloids with potential harmful effects on humans and the environment. This group consists of metals elements such as cadmium, lead, copper, mercury, nickel, zinc and chromium. Although some of these metals are important for human biochemical processes in low concentrations, such as zinc which is a cofactor for several enzymatic reactions, greater concentrations can cause morbidity or mortality. Others are only toxic to humans, even in small exposure, such as lead, cadmium and mercury (Soghoian 2009). Identification and removal of the source is vital to stop the exposure.

While heavy metals are elements naturally occurring in the earth's crust, most of the human exposure and environmental contamination is a result of activities such as mining, domestic and agricultural use of metals and metal containing compounds and industrial production and use. Industrial sources include coal burning in power plants, metal processing in refineries, petroleum combustion, plastics, textiles and microelectronics (Tchounwou et al. 2012).

3.2 Health Effects

Several major studies have established the relation between exposure to air-borne pollutants and hospital visits and premature death (Sarnat et al. 2001; Greven et al. 2011). The strongest and most consistent relation found is for particulate matter. The Adventist Health study of Smog (AHSMOG) in the US found significant effects on non-malignant respiratory deaths in men and women, and lung-cancer mortality in male non-smoking Seventh-Day Adventists, due to exposure to particle matter with a diameter less than 10 micrometers (PM10) (Abbey et al. 1999). Other studies have shown an association between increased cardiovascular morbidity and mortality for elderly people and their episodic exposure to particle matter (Gordon & Reibman 2000). The heart rate of retirement home residents were shown to have strong correlation to the variability of daily PM levels, and in addition, plasma viscosity changes has been associated with the levels of ambient PM. These effects are dependent on both the concentrations and length of exposure and long-term exposure have cumulative effects that is larger and more persistent than short-term exposure (Pope 2007).

3.2.1 Gaseous pollutant effects

Several WHO[25] studies conducted in 2000, 2006, 2013 and 2014 has examined the harmful effects of gaseous pollutants and particle emissions.

Nitrogen oxides

The levels of nitrogen oxide and nitrogen dioxide has risen due to combustion of fossil fuels, primarily in diesel motor vehicles, but also from stationary sources such as heating and power generation. Exposure to NO_2 can be dangerous to children, as it may increase the symptoms of bronchitis for asthmatics, and reduce the growth of the lung function.

Ozone

High levels of ground-level ozone is associated with breathing problems, reduced lung function, raised number of asthma attacks and other lung diseases.

[25] http://www.who.int/mediacentre/factsheets/fs313/en/

Sulphur oxide

Elevated levels of sulphur oxide can affect the function of the respiratory system and cause eye irritation.

Carbon monoxide

Carbon monoxide inhibits oxygen uptake, and can be harmful to humans in high concentrations.

3.2.2 Particle matter

Particle matter (PM) in its various forms has several severe effects on human health. The small size of the particles enables them to penetrate the thoracic region of the respiratory system. The PM10 particles is proven to have the most severe effects from short-term exposure, while PM2.5 has stronger association with the effects of long time exposure. (World Health Organization 2013).

There are several respiratory effects associated with exposure to PM. PM has been clearly identified as a factor for reduced lung function and exacerbation of asthma. In addition, it has cardiovascular effects, such as increasing the risk of myocardial ischemia and endothelial vasomotor dysfunction.

Short-term exposure to PM

There has been abundant studies on the short-term health effects of PM, with focus on hospital admission and mortality (Brunekreef & Holgate 2002). From APHEA-2, an European study covering more than 43 million people living in 29 cities, shoved that for each 10 ug/m3 increase in PM10, the all-cause daily mortality increased by 0.6%. Hospital admissions of people suffering from cardiovascular disease (CVD) increased by 0.5%. People over 65 suffering from asthma and chronic obstructive pulmonary disease (COPD) the admissions were increased by 1.0% per 10 ug/m3 PM10. In another study covering 1.843 million people over the age of 65, the effects of PM10 were a 1.5% increase in COPD admissions and a 1.1% increase in CDV admissions per 10 ug/m3 of PM10.

Long-term exposure to PM

For long-term exposure, meaning exposure lasting for months and years, PM2.5 is proven to be the most harmful. Long-term exposure is associated with an increase in the risk of cardiopulmonary mortality by 6-13% per 10 ug/m3 of PM2.5 (World Health Organization 2013).

3.2.3 Sensitive groups

Certain population groups are more vulnerable than the general population to the effects of air pollution. Children, elderly and people with sensitive airways have respiratory systems more susceptible to air pollution. Since the exposure is ubiquitous and involuntary, sensitive groups might benefit from air quality information to evaluate their daily needs against the cost of their exposure.

Children and newborns

Newborns and small children are in the process of developing their airway systems, and therefore more susceptible to air pollutants. There has been reported that increased exposure to traffic pollution and elevated ozone levels is associated with increased risk of asthma development and reduced lung function (Bernstein et al. 2004; Peters et al. 2014).

Elderly

In comparison to the general public, elderly people have increased susceptibility of air pollution due to normal and pathological aging and related processes. The predominant chronic diseases are cancer and cardiopulmonary diseases. In addition, chronic obstructive pulmonary disease (COPD) and asthma increase the decline of lung functions and frequency of mortality. (Bentayeb et al. 2012).

Diseases

People suffering from respiratory and cardiovascular diseases have weakened health that make them particularly susceptible to air pollution.

Air pollution affect people with a number of chronic respiratory diseases, such as asthma, bronchitis, Chronic Obstructive Pulmonary Disease (COPD) and allergies[26].

There is also discovered relations between air pollution and cardiovascular diseases, such as heart attacks and stroke, and to lung cancer.

3.2.4 Informing the public

As explored previously in this chapter, the links between different air pollutants and health effects is difficult to assess, and require expertise from several research fields employing various types of studies and models. Aoki et al. (2009 p. 8) states "*links between specific pollution sources and health are often difficult to establish, remediation is often complicated and costly, and parties often have competing interests*."

To raise the awareness of air pollution, both to the general public and to people with higher susceptibility for air pollution, an Air Quality Index has been developed.

Air Quality Index

Air Quality Index (AQI) is a scale developed to communicate the measured levels of air pollution in a given area to the public. Although the numbers of pollution concentrations in various scales are important for research and air quality modeling, the public might not be aware of what constitute as harmful levels. A simplified scale using colors can better convey the health risks associated with the exposure to the levels of air pollution measured and modeled.

NILU, with more than 40 years of experience in air quality measurements and modeling, provides a publicly available Air Quality Index for Norway. This is available at the website luftkvalitet.info. The site provides a map of the measuring locations in Norway, along with color codes for the locations, indicating the level of air pollution in general. The color scale go from green indicating low pollution, through yellow for some air pollution, orange for high pollution, and red for alerting of very high pollution levels. If there is no data on a location, the color is white. These air pollution indications relate to the locations of the

[26] http://www.who.int/mediacentre/factsheets/fs313/en/

monitoring stations providing the data, and is unable to provide any detailed assessment of the local air pollution.

3.3 Meteorological effects

Studies done in urban environments have shown that variances in meteorological factors affects the concentration and distribution of different air pollutants. Thai, McKendry, & Brauer (2008) has studied how different meteorological factors affects the types of PM.

Temperature

Elevation in temperatures may affect both the concentrations of PM2.5 and ultrafine particles. Due to increased coagulation of the ultrafine particles into PM2.5 in higher temperatures, the PM2.5 concentration increase as the concentration of ultrafine particles decrease.

Wind speed

Increased wind speeds might dilute emissions from primary sources of PM. It may also re-deploy PM10 particles from the atmosphere to the ground. There has also been discovered a negative correlation between wind speed and concentrations of ultrafine particles, PM2.5 and CO (Jarjour et al. 2013).

Precipitation and humidity

Precipitation in the form of rain has a negative correlation with both PM2.5 and PM10. This can be explained with increased dilution due to the atmospheric conditions along with low deployment of particle matter present in the ground (Thai et al. 2008). Ultrafine particles are not similarly affected since they are emitted directly to the atmosphere from combustion, do not get affected by washout, and has a short lifetime before they coagulates into larger forms of particles.

Inversion layer

This is a meteorological phenomenon where a layer of warmer and less dense air acts as cap for the colder air below and prevents normal air circulation (Seaton et al. 1995). This inhibits transportation of the pollutants, and thus leading to elevated air pollution levels. This phenomenon occurs in many big cities, such as Los Angeles, Mexico City and Mumbai, but also in the smaller Norwegian cities of Oslo and Bergen.

3.4 Urban effects

In the urban environment, several effects affect the concentrations and dilution of pollutants. The most important is the amount of traffic and its speed.

Traffic density and speed

Both the traffic density and the speed of the vehicles highly affect the levels of pollutants emitted to the environment (Hertel et al. 2008). The levels of pollution are clearly highest during the morning rush on weekdays, due to rush traffic.

Concentration peaks

Peaks of pollution concentrations is normal in the urban environment. These concentration peaks come from a variety of possible sources, such as from cycling behind a bus or large vehicle with high emissions, or by passing construction sites (Thai et al. 2008) or other non-traffic sources such as restaurants(Peters et al. 2014).

Extraordinary events

Fires, accidents and other events happen in the urban environment from time to time and can result in hot-spots of pollutants over extended periods.

3.5 Transport modes

The choice of transport mode and its impact on the personal exposure has been researched in several studies (Zuurbier et al. 2010; Dons et al. 2012; de Nazelle et al. 2012). The personal exposure in traffic is very high, and various factors that affect the exposure is:

Respiration

The rate of respiration varies due to the level of activity, and is higher during active transport modes such as walking or cycling. The level of respiration at active travelling might increase to 6-7 times higher than compared to levels during sleep (Dons et al. 2012).

Route selection

Several studies (de Nazelle et al. 2012; Dons et al. 2012; Peters et al. 2014) has concluded with the fact that the selection of low-traffic routes and bicycle paths separated from traffic can decrease the exposure to air pollution and potentially reduce their harmful effects on the health.

Air circulation and fuel type

The circulation of air is not a concern for people walking or cycling, but in closed vehicles in traffic such as cars and buses, there has been found elevated levels compared to the open travel modes, particularly for travels by diesel-fueled buses but also in private cars (Zuurbier et al. 2010).

3.6 Diurnal and weekly variations

The diurnal variations in the urban air pollution is closely linked to the traffic activity (Zwack et al. 2011; Peters et al. 2014; de Nazelle et al. 2012). Levels of air pollution is normally at the highest during the morning commute, lowering during the day, rising again in afternoon rush hours, and declining in the evening and night. During weekends, the pattern is the same but at lower levels, due to lower numbers and more even patterns of traffic.

3.7 Conclusion and findings

Air pollution and health effects are a complex multidisciplinary field requiring both short-term and long-term research. Due to this complexity, this thesis has merely grasped the topics, but still the main conclusion is that various air pollutants in the urban environment can contribute to severe effects on the human health, especially for infants and children, elderly and others with various cardiovascular and respiratory diseases.

A diverse set of external affects the air pollutants in the urban environment, changing their composition and distribution. Air pollution modeling use this knowledge to increase their performance and accuracy in the predictions. Such modeling will be explored in the next chapter, and both traditional models and mobile monitoring based models will be reviewed.

4 Air Quality Modeling

This chapter address the field of air quality modeling, both the traditional models used in environmental research today, and a possible new model for creating real-time models based on mobile monitoring data.

4.1 Models for static air monitoring

To assess the pollution levels and predict future concentrations, researchers and regulators employ a number of different air quality models. Depending on the type of model and desired outcome, they take a set of parameters and mathematically calculate the pollution levels. The models simulate physical and chemical processes affecting air pollutants and their reaction and dispersion in the atmosphere[27]. The input parameters for the models can be emission rates and meteorological data, which estimates the emission of primary pollutants, and for some models, secondary pollutants formed by atmospheric chemical reactions. The models are important for air quality assessment, and is used to identify source contributions of pollutants and develop strategies to reduce the harmful emissions. Additionally, they can assess whether a new emission source will exceed the accepted air quality levels, and predict future pollutant concentrations from multiple sources.

The most common models can be split into three groups:

Dispersion modeling
Dispersion models focuses on estimating the concentrations of pollutants at specified ground-level receptors surrounding an emission source.

Photochemical modeling
Simulate the impact from all sources by estimating the concentrations of pollutants and deposition of both inert and chemically reactive pollutants over large spatial scales.

Receptor modeling
Receptor models are observational techniques using the chemical and physical

[27] http://www.epa.gov/scram001/aqmindex.htm

characteristics of gases and particles measured at source and receptor to both identify the presence of and to quantify source contributions to receptor concentrations.

4.2 Models for urban air quality monitoring

To be able to model the air quality in urban environment there is a range of different modeling techniques applied. Urban features, such as traffic density and building height, affects the atmospheric flow and microclimate. Srivastava & Rao (2002) describe various models used for urban air quality monitoring:

Eulerian models (grid models)

The Eulerian models predict concentrations of air pollution in urban areas. The space domain or location are divided into grid cells. For each cell the transport, diffusion, transformation and deposition of pollutants are described by a set of mathematical expressions.

Examples are of such models are:
CALGRID[28] - a photochemical grid model developed for regulatory assessment of ozone control.

FARM[29] – Grid model for dispersion, transformation and deposition of reactive photochemical and particulate pollutants.

DEHM[30] (Danish eulerian hemispheric model) – A collection of models developed to provide realistic description of the physical and chemical processes in the atmosphere.

Lagrangian models

The Lagrangian models focuses on calculating wind trajectories, and transportation of air along the trajectories. . The model assumes an even distribution of pollutants within the boundary layer, and considers only a simplified exchange within the troposphere. As the model performs computation on a smaller number of moving cells opposed to the eulerian

[28] http://www.arb.ca.gov/eos/soft.htm
[29] http://www.aria-net.it/front/ENG/codes/files/10.pdf
[30] http://envs.au.dk/videnudveksling/luft/model/dehm/

models that performs calculations on each cell, the computational cost of modeling is much lower than on eulerian models.

In addition to these two, other models for urban air quality assessment are available:

Box models
The whole domain is considered as box, into which emitted pollutants undergo chemical and physical processes. The model is not able to provide any local concentration information since the whole domain is included in the box.

Receptor models
Each source of contribution, or groups of sources, is apportioned to the measured concentrations, without considering the dispersion patterns of the pollutants.

Computational fluid dynamic models
The models calculate and analyze the flows in urban infrastructure and areas.

Gaussian steady-state dispersion models
The models focuses on plumes of emission, and calculate the impact of plumes at the maximum ground level and the distance of maximum impact of the source.

Model requirements

Srivastava & Rao (2002) also states that the models for urban air quality requires a number of input parameters.

Sufficient ambient air concentration monitoring data.
The geography of the urban area, construction activity, road network, location of buildings etc. all interplay in the dispersion of pollutants. To understand the ambient status of pollutants, it is necessary to have a strategically distributed network of monitoring locations.

Micro meteorological data
Different meteorological conditions such as temperature, humidity and wind direction changes in the urban environment due to anthropogenic activities and architecture.

Knowledge of all sources of emissions

Both the knowledge of the sources of emissions and their emission profiles are required for the predicative modeling.

Atmospheric chemistry

The transformations of emitted chemical species and their reaction rate pathways must be known to model for observed concentration of pollutants.

Health impacts

The models should be able to incorporate the health effects of the urban air pollution.

None of the models available is able to handle all the requirements for urban air quality modeling alone. As each are specialized on one certain aspect, utilization of several models are required for accurately model the urban air quality domain.

AirQUIS Dispersion models[31]

AirQUIS is a commercial modeling engine developed and offered by Norsk Institutt for Luftforskning (NILU). An extensive set of features is offered, such as data acquisition, visualizations, reporting as well as a range of models for air pollution estimations.

EPISODE[32]

Episode is a numerical dispersion model developed by NILU, as a part of the AirQUIS suite.

A part of the model is eulerian, and produces a three-dimensional grid of the numerical solution of the atmospheric conservation equation of the pollutants.

The model also has two separate Lagrangian sub-grid models, one for line sources and one for point sources. The sub-grids is used specifically to calculate spatial emission concentration distributions that are close to main roads and individual point sources on a sub-grid scale.

---

[31] http://www.nilu.no/airquis/models_dispersion.htm

[32] http://www.nilu.no/airquis/models_episode_long.htm

The horizontal resolution of the grid models adjusts to preference, and while the normal grid resolution is 1km, the grid is adjustable to smaller scales such as 50-100 meters.

4.3 Models for mobile monitoring

The new paradigm of mobile monitoring is able to produce spatially distinct data with a high resolution. The mobile monitoring extends the data already available from the fixed-site monitoring stations, and provides a level of detail not available with the traditional monitoring and modeling techniques (Allemann & Raubal 2013). The new detailed maps of the spatial and temporal variations in urban environments calls for new methods of analyzing and modeling of the data. The OpenSense mobile monitoring project described earlier has researched how this real-time data can assess the levels of pollution in the urban environment, and the exposure to the inhabitants.

OpenSense

The OpenSense project has developed land-use regression (LUR) models to assess the interurban concentrations of pollutants (Hasenfratz et al. 2014). These models use variables such as land-use and traffic density to model concentrations of pollution at locations not covered by mobile sensor nodes. The project chose LUR models over other modeling techniques, such as proximity-based assessment, statistical interpolation and line dispersion models, mainly due to the relatively low computational overhead. This is highly beneficial for producing a large number of models. In addition, the LUR models has historically been used to predict a wide range of air pollutants. The models do not consider meteorological factors, as the study state that meteorological conditions have to be known for each measurement location, and there has to be significantly different meteorological conditions among the locations at the given time for the integration to be successful.

Analyses of the model show that the model is able to produce good quality maps with a high spatial resolution of 100m * 100m grid on yearly to weekly time scales, but struggles with sub-weekly temporal resolutions. The researchers propose a solution that uses historically data of pollution concentration with similar meteorological conditions to improve the quality of the model.

Other projects

The two other community sensing projects describe earlier have not published any papers regarding models for creating high-resolution maps of the urban pollution. Peters et al. (2014) do not use grid-modeling techniques, but instead map the concentrations on the routes with 10m spatial resolution.

4.4 Mobile monitoring models vs traditional models

Mobile monitoring and modeling is radically different to the traditional way of monitoring and modeling of air pollution. The table below describe the main differences found.

	Mobile monitoring models	Traditional models
External data input	Low number of additional parameters needed	A high number of additional parameters from a variety of sources is needed
Peak events	Peaks are registered due to the high data resolution	The measurements are averaged and unable to register the peaks
Equipment cost	Possibility for utilizing low-cost equipment for mobile data gathering	Professional equipment placed at strategic monitoring locations
Map resolution	Very high resolution possible, depending on the number of nodes and their resolution	The low number of nodes increase the need for accurate modeling
Reliability	Still in the inception phase, needs more research to verify the accuracy and quality	Models tested and verified by the scientific community, generally high reliability

Table 1 - Mobile monitoring vs traditional monitoring

Aoki et al. (2009) has done a qualitative analysis of environmental monitoring, investigating both traditional modeling and the new paradigm of real-time mobile monitoring. From interviews of government, public and private stakeholders, the researchers provides insight into how real-time sensing can fit into the context of environmental action, and how the stakeholders can make sense of the data produced.

One of the key points derived from the analysis is the question of data quality, and if the data obtained from mobile monitoring good enough to be used for scientific purposes. Although debate on whether the traditional air quality monitoring is able to provide an accurate picture of the air quality, and real-time high-resolution monitoring is considered of great inherent value, the question is how reliant are the data gathered and how can this data be used.

In the survey there was raised concerns about the political relevance of the real-time high-resolution data, the quality of the data obtained from such endeavors. Also there was disagreement on the value of citizen participation, on their incentives to engage themselves and the credibility of their methods.

4.5 Exposure modeling

In order to assess the personal exposure from air pollution, there are models that calculates the exposure based on the spatial and temporal concentration measurements and the time spent at the monitored location.

Several studies look into the personal exposure, (MacNaughton et al. 2014) claims that the type of bike route the cyclist choose has a significant impact on their personal exposure, but no health exposure assessment is done.

4.5.1 Grid modeling of exposure

Peters et al. (2014 p. 41) state that "*exposure of a cyclist is assessed by integrating the spatio-temporal concentration measurements in time*." They also describe a formula to calculate the exposure as follows: "*Exposure of a cyclist is hereby defined as (Klepis, 2006):*

where Ci(t, x, y) is the concentration occurring at time t at location (x, y) and t1 and t2 the starting and ending time of the exposure episode."

$$E_i = \int_{t_1}^{t_2} C_i(t, x, y) \mathrm{d}t$$

By using this formula on a grid model, as the one proposed for the real-time data modeling, one can assess the exposure for the single grid cells. The calculated exposure for each cells can be summarized to provide the total exposure for the route.

4.5.2 Hot-spot and peak exposure events

As seen in studies discussed earlier (Peters et al. 2014; Pattinson et al. 2014), the occurrence of hot-spot and peak events can have a significant impact on the integral exposure obtained and the effects on health. These events, although important for the exposure assessment, are not included in the traditional models due to the low spatial resolution. Real-time monitoring with high spatial and temporal resolution is able to provide such data, thus able to include the occurring events. Because of this, the real-time data is able to provide a more realistic exposure assessment than the traditional models.

5 Mobile Monitoring

This chapter presents a literature study on five mobile monitoring campaigns that used bicycles as the transport vehicle. The methods of research and the monitoring equipment is reviewed, the study conclusions and issues, and finishes with a summary and recommendations for the campaign.

5.1 Literature review of mobile monitoring campaigns

Urban air quality monitoring as a research topic has been executed in a number of ways. The traditional way is to use fixed location measurement stations. These stations measure the air pollution levels at strategic points. Together with data on other confounders such as meteorological conditions and traffic intensity, various models process air quality data to provide an assessment of the distribution and intensity of air pollutants. As the measuring environment is stable and the equipment is of professional grade, the data provided are of high quality and reliability. It is also able to assess the temporal variations in the air pollution. However, due to their stationary nature, they are unable to measure the spatial variances in the air pollution.

Mobile monitoring is able to provide data with high spatial resolution, and can thus help assess variations in the distribution of air pollution. In addition, mobile monitoring is able to add the temporal dimension, giving the possibility to monitor the changes in air quality over both space and time. As the urban environment has high variability in the concentrations of air pollution, the combination of spatial and temporal data has several benefits over the traditional stationary monitoring (Van den Bossche et al. 2014).

Various portable equipment for mobile monitoring are increasingly utilized in studies on exposure to air pollution. Two major tracks for exposure assessment using mobile monitoring is found in literature. Most of the studies applies personal monitoring with portable sampling equipment or real-time monitors equipped on study subjects (Peters et al. 2014). They address the personal exposures in urban environments, for example special groups such as children with asthma, or the difference in modes of travelling, such as car, bicycle and public transport.

The other track of studies address the potential of using the mobile measurements to generate maps of the air pollution. The addition of spatial data from mobile monitoring to the existing temporal data available from the fixed sites enables improved estimates of the air pollution on a high-resolution scale. The scientific community can use the data to improve the ambient modeling of air pollution variability and distribution, exposure estimates and health effect studies. The general public will have the benefit of the improved resolution as the data enables the production of near real-time air quality maps (Snyder et al. 2013). Providing visualization of the data through smart phone applications enables people to use the data to assess their individual personal exposure to air pollution.

The research field of using low-cost monitoring equipment in mobile campaigns has until now been held back due to low availability of reliable sensor equipment. There has been studies exploring the temporal and spatial variability of air pollution in urban environments, but these studies use expensive professional-grade portable equipment for the monitoring.

Albeit the lack of studies using low-cost sensor platforms on bicycles, methods applied in the other campaigns can be studied to create a campaign methodology for the low-cost sensors building on the best practices found. The community sensing projects described in chapter 1 has produced several studies that examine how to plan a campaign, and the amount of data that needed to create real-time mapping of the air pollution levels.

5.1.1 Mobile monitoring methods

The studies selected for the best practices review share the treats of being oriented to mobile monitoring, and with the goal of investigating the spatial and temporal variations in the urban environment. They are also situated in cities located in western parts of the world that has urban environments and levels of pollution similar to Oslo.

The studies are published in the journals Atmospheric Environment and Science of the Total Environment, which are well renowned journals in the Environmental Science field.

Duration

The monitoring campaigns is often very limited in time, and is thus only able to provide a snapshot of the pollution levels (Van den Bossche et al. 2014). The campaign with the longest duration found, a campaign from the OpenSense project, use a novel monitoring platform mounted on public transport, and the data collection span for two years (Hasenfratz et al. 2014). Other campaigns with bicycles as the transport vehicle (Peters et al. 2014) (Pattinson et al. 2014) (MacNaughton et al. 2014) (Thai et al. 2008), the duration were limited to weeks and months. The limitations in duration make the campaign data obtained only representative for days of similar weather conditions and seasons.

Location

Most of the studies explore the notion of urban air pollution, as the urban environment both inhabits the largest number of people along with the highest traffic densities. The selected monitoring campaigns has taken place in cities of Western Europe (Peters et al. 2014) (Hasenfratz et al. 2014), Vancouver in Canada (Thai et al. 2008), Boston in USA (MacNaughton et al. 2014) and Auckland area of New Zealand (Pattinson et al. 2014). These cities share the treats of approximately the same climatic zone and level of urban development and life standards, and can thus be comparable to some extent.

Monitoring area and runs conducted

All of the monitoring campaigns with a bicycle as the transport vehicle follow certain predefined routes. These routes encompass different sets of urban microenvironments to investigate the spatial distribution and variation of air pollutants.

The table below summarizes the different bicycle based studies on their number of predefined routes, the length of the routes, and the number of monitoring runs for each route. The timeslots for monitoring and their length in time show the diurnal variations in monitoring. The number of reference stations refer to traditional fixed-site monitoring stations used to obtain reference values of the pollution levels. The duration show the length and season of the monitoring campaigns.

	Peters et al.	Pattinson et al.	Thai et al.	MacNaughton et al.
Routes	2	2	1	5
Length	2km + 5km	15,9km + 19km	20 km	unspecified
Runs	258 + 96	20	15	20
Timeslots	07:00 – 13:00	07:00 12:00 17:00 22:00	07:00 – 09:00	07:00 - 10:00 15:00 - 18:00
Reference stations	1	1	4	unspecified
Duration	Feb - Mar 2012	May - July 2010	Aug - Oct 2007	unspecified

Table 2 - Campaign route summary

The mobile monitoring campaign with the largest coverage was done by Peters et al. (2014) in Antwerp, Belgium. In two months, the campaign gathered data over two separate routes, totaling in 354 separate runs covering a total of 996 km in length. The study by Pattinson et al. (2014) covers 698 km in total for both routes, Thai et al. (2008) covers 300km from 15 repetitions of a 20 km route, and MacNaughton et al. (2014)

The Antwerp campaign had a preference of many repetitions of a smaller area in contrast to the other campaigns that covered larger areas, to increase the resolution and reliability of the data obtained (Van den Bossche et al. 2014).

Air Monitor Equipment and pollutants monitored

The equipment used in the bicycle based mobile monitoring campaigns was to a large degree focused on monitoring the exposure to black carbon and particulate matter of variable sizes. This was monitored by professional grade portable equipment that was carried along in a backpack or in a cart pulled by the bicycle.

The temporal resolution was in most cases set to the lowest possible setting that still allowed for valid data collection. The lowest resolution was at one-second intervals, and the highest was at 30-second intervals.

Additionally, all of the studies used some form of GPS device to collect the coordinates for the campaign data, some collected meteorological data used for comparisons and analysis, and some used a camera to record video of the campaign.

The table below describe the different pollutants monitored in the campaigns, the equipment used to monitor the pollutants, and the temporal resolution of the monitoring.

Pollutant	Peters		Pattinson	
	Equipment	Resolution	Equipment	Resolution
UFP	TSI model 8525 P-trak	1 sec	TSI model 3007	1 sec
PM 2.5				
PM10			GRIMM 1.107	6 sec
BC	microAeth Model AE 51	1 sec		
NO_2				
O_3				
CO_2				
CO			Langan Products Model T15n	1 sec

Pollutant	Thai		MacNaughton	
	Equipment	Resolution	Equipment	Resolution
UFP	TSI model 8525 P-trak	10 sec		
PM 2.5				
PM10	GRIMM 1.108	10 sec		
BC			microAeth Model AE 51	1 sec
NO_2			CAPS NO_2 monitor	1 sec – 1min averaged
O_3				
CO_2			TSI model 7565 Q-trak	30s
CO			TSI model 7565 Q-trak	30s

Table 3 - Equipment comparison

5.2 Recommendations for measurements

The studies have investigated how to perform monitoring campaigns on bicycles, and provides conclusions that to a large degree correspond to each other.

5.2.1 Conclusions

The studies have addressed both the spatial resolution possible to obtain from mobile monitoring campaigns, and the possibility to reduce the personal impact of the pollution by choosing alternative less polluted routes.

Spatial resolution

Several of the projects (Peters et al. 2014; MacNaughton et al. 2014; Thai et al. 2008) discovered that it is possible to obtain datasets with a spatial resolution down to street-to-street levels and intra-street levels, and have found considerable variations in the pollution on these levels.

Exposure reductions

The studies by (MacNaughton et al. 2014; Thai et al. 2008; Peters et al. 2014) states that the choice of route and cycling path has considerable impact on the personal exposure to air pollution. Choosing a route with less traffic, or a bicycle path separated from the traffic, can give a high reduction from the negative health impacts of air pollution.

5.2.2 Open issues

The aforementioned studies have their limitations and issues, such as the types of monitoring equipment and the area and time constraints of the campaigns.

Low-cost sensors

The utilization of low-cost sensors is only found in the study by Hasenfratz et al. (2014), all the other studies perform the measurements with expensive monitoring equipment specifically developed for environmental monitoring, tested and verified for such use.

Time span of campaigns

The relative short time-spans of the bicycle campaigns sets limitations on the datasets provided, thus only providing a snapshot of the situations for the campaign locations (Van den Bossche et al. 2014).

5.2.3 Summary

Traditional air quality monitoring and modeling is based on a low number of fixed-site monitoring stations, and thus are the additional data required by the models of high importance for the quality of the output.

The new paradigm of mobile monitoring tilts the situation around, as the number of mobile monitoring units can be greatly increased and thus provide a high-resolution map with real-time or “near real-time” details on the levels of air pollution. This eliminates much of the need for additional data, but as there can be gaps in the distribution of the monitoring units, a lightweight linear regression model as the model from the OpenSense project (Hasenfratz et al. 2014) can be used to fill the gaps.

Van den Bossche et al. (2014) has done a study to validate the methodology used in the study by (Peters et al. 2014). This is also the study with the largest dataset from the literature findings. They state that to create high-resolution maps many repeated runs are necessary, and one should favor more repetitions of a short route than longer runs with less repetitions. The campaigns should be carefully set-up with a sufficient number of repetitions of fixed routes in relation to the data quality and the spatial resolution needed. Additionally, the campaign hours should be uniform to increase the size of the datasets for these hours.

Additionally, the study by (Peters et al. 2014) describe a multiscale approach to the analysis phase. By refining the temporal and spatial resolution into different entities, the whole routes, different streets and down to points within streets could be examined for spatial variance, and refining into in days, hours and seconds for temporal variance.

The next chapter will describe a monitoring campaign done in the spring of 2015 at Majorstua in Oslo, utilizing the monitoring platform developed for the Citi-Sense MOB project.

6 Mobile Monitoring campaign

As found in the literature study of mobile monitoring, street level spatial variations in urban air quality has been discovered in addition to the temporal and diurnal variations already known from fixed-site monitoring. These variations is found both between streets, and in different segments of the same street.

Much of the research uses professional grade equipment with a high price tag, and this limits the possibility for larger monitoring campaigns. The campaigns has also to a large degree focused on particulate matter in various sizes or black carbon emissions.

The availability of the lower-cost sensors for various gaseous air pollutants enable a new type of mobile monitoring where the cost of the sensor equipment has been radically reduced. The lower cost of the monitoring equipment facilitates larger studies, as many more monitoring platforms can be produced for the same cost as the professional grade equipment. As there is little available studies on the spatial distribution of air pollutants such as NO_2 and O_3, the use of the new sensors in such campaigns has the ability to provide new insight into the fields of both mobile monitoring as well as the distribution of these pollutants in an urban environment.

The Citi-Sense MOB platform will be the monitoring platform for this study. The platform detects levels of several air pollutants, along with meteorological factors, time and location coordinates. The platform is mounted on an electric bike, which provides the operational power for the platform.

This methodology driven campaign will try to assess the temporal and spatial variations of air pollutants in the urban environment of Majorstua in Oslo. Majorstua is an inner city area consisting mainly of low to medium rise residential buildings, some with commercial areas on the ground floors. Two separate routes is selected for this campaign. The first one includes an arterial road of the ring 2 road network, a low trafficked residential street near a park, and an inner city shopping area street. The second route is located inside the first route, and follows the parallel streets of the Fagerborg residential area of Majorstua.

6.1 Campaign questions and challenges

As discovered by the literature review, the notion of utilizing low-cost sensors for bicycle mobile monitoring campaigns is new in the field of air quality monitoring. Therefore, in addition to investigating the spatial and temporal variations of air quality in the urban environment, the study will also review the functionality and quality of the monitoring equipment and the quality of the data obtained.

6.1.1 Spatial and temporal air quality variations

The spatial and temporal variation of air pollution has been identified in several studies, even down to intra-street levels. As these campaigns utilize traditional high cost professional equipment, there is little knowledge of using the low-cost sensors for the same purposes. This thesis will conduct a campaign in Oslo to monitor the air quality with the low-cost sensor platform. The campaign seeks to gather knowledge on the temporal and spatial variability of gaseous pollutants in the urban environment, and the impact of choosing low traffic roads instead of the main roads.

6.1.2 Mobile monitoring data quality

Because of the lack of other results from using the low-cost sensors, the stability and quality of the measurements obtained in the campaign needs verification. Reference values from the fixed-site monitoring in Kirkeveien will be used in this comparison, providing insight into how the mobile monitoring data stand against reference values.

6.1.3 Mobile monitoring equipment

The monitoring equipment used in this campaign is developed specifically for the Citi-Sense MOB project. As this project has not previously done any campaign with the platform, a review of the experiences from the campaign is performed to give feedback on the status of the equipment.

6.2 Monitoring platform

The Citi Sense MOB project aims to extend the already produced knowledge on air quality from stationary monitoring units by adding data addressing the temporal and spatial varieties in air pollution levels. To achieve this goal, the project utilizes various mobile sensor units, ranging from small units that people are able to carry without much effort, to larger sensory units placed on bicycles and public transportation. These sensor units for the person and bicycle measurements are of relatively low cost. This make it affordable to utilize a substantially larger number of sensor devices, thus greatly increasing the amount of data available. In addition, this lower the adoption cost for external contributors, and can boost efforts in recruiting additional attributers.

For the selection of monitoring platform, the Citi-Sense MOB project did an assessment of 15 different platforms, and five of these were analyzed in detail. The sensor platform chosen for the project was the first choice primarily due to the availability and communication capabilities, but also its size and expected functionality was important for the decision.

6.2.1 The DunavNet platform

The sensors and control board are a situated under a sensor platform, developed by DunavNet[33] as a pre-commercial platform for the Citi-Sense project. It enables measurements of several gaseous air pollutants, meteorological parameters and context data such as location and time. The platform records measurements every second, and averages them to 1-minute measurement averages. The 1-minute average values are sent via GPRS to a central database.

Platform characteristics

- Dimensions: 225 * 150 * 100 mm
- Weight: Approx. 500 gr
- Ambient temp range: - 20 to +50 °C
- Voltage supply range:
    - For auto industry 8-28V DC

[33] www.dunavnet.eu

        - For Fixed/indoor 12V DC
- Power consumption max 10W
- Configuration update via SMS
- TCP/IP or UDP/IP protocol support
- Data visualization by web widgets

6.2.2 Sensors mounted on the platform

The sensors mounted to the platform focuses primarily on measuring the different levels of gases present in the atmosphere, but there are also additional sensors that measure the temperature, the relative humidity and the atmospheric pressure.

The sensory board are equipped with the following sensors:

- Sensor for atmospheric pressure: **MPX4115** (15 - 115 kPa, 46 mV/kPa)
- Sensor for temperature and humidity: **SHT71** (-40°C – 123°C, 0-100%)
- Sensors for concentrations of air pollutants (Alphasense[34])
    - **Carbon monoxide - CO-B4**
      (0-1000ppm, 30 to 55 nA/ppm in 400ppm CO, -30°C - 50°C, 80 - 120 kPa)
    - **Carbon dioxide - CO_2-IRCA1**
        - (0-5000ppm, -30°C(P/T ratio 7%) – 50°C, (NTC, R25 = 3000 Ω B= 3450 K))
    - **Nitrogen monoxide - NO-B4**
        - (0-100ppm, 550 to 850 nA/ppm in 400ppm NO, -30°C - 50°C, 80 - 120 kPa)
    - **Nitrogen dioxide - NO_2-B4**
        - (0-20ppm, -200 to -450 nA/ppm in 1000ppm NO_2, -20°C - 50°C, 80 - 120 kPa)
    - **Ozone - O_3-B4**
        - (0-2ppm, -850 to -1700 nA/ppm in 1000ppm O_3, -20°C - 50°C, 80 - 120 kPa)

Communication technologies

The platform has a GPRS communication module that, via a SIM card using public mobile networks, are able to report the various readings to a central database on a given time interval.

[34] http://www.alphasense.com/

GPS

To track the geographical position of the bike the platform is equipped with a GPS unit, which adds the latitude and longitude of the location measured to the sensory data provided from the platform.

6.2.3 The electric bike

As the DunavNet platform needs a constant power supply of 12V an electric bike became the obvious choice for the project. Main criterions for the selection of the bike was access to the power supply from a separate regulator, app connectivity and collaboration with the producer. The chosen bike has a separate controller with a 12V cable that can power the platform. The brand of the bike is BH Emotion[35], delivered to the project from EVO[36] Oslo.

6.2.4 Protective casing

To protect the sensor, electronics and connections of the platform, it is installed in a plastic protective container. Inside the container Styrofoam pads surround the platform to keep it in place and to protect it from vibrations. Based on considerations on exposure to dirt and water, airflow to the sensors and power availability, the container is mounted on the front part of the baggage board on the bike. It has an opening underneath that exposes the sensors to the flow of air, enabling them to do the measurements.

6.2.5 Reference monitoring station

In Schwachs street, approximately 10 meters from Kirkeveien, there is a fixed-site monitoring station operated by NILU that gathers data on the levels of air pollution in the area. This station operates continuously, and are thus able to assess the temporal variances of air pollutants for the site.

The data from this station have a high reliability as the monitoring equipment is of professional grade, and calibrated regularly to ensure the validity of the data. The monitoring data contains data on gaseous pollutants such as NO_2 and O_3 on an hourly

[35] http://www.bhbikes.com/web/en/ebikes/evo.html
[36] www.evosykler.no

average, in addition to data on particulate matter. Both the location and the high reliability of the data from the fixed-site monitoring enables the use of the data as reference data for the campaign.

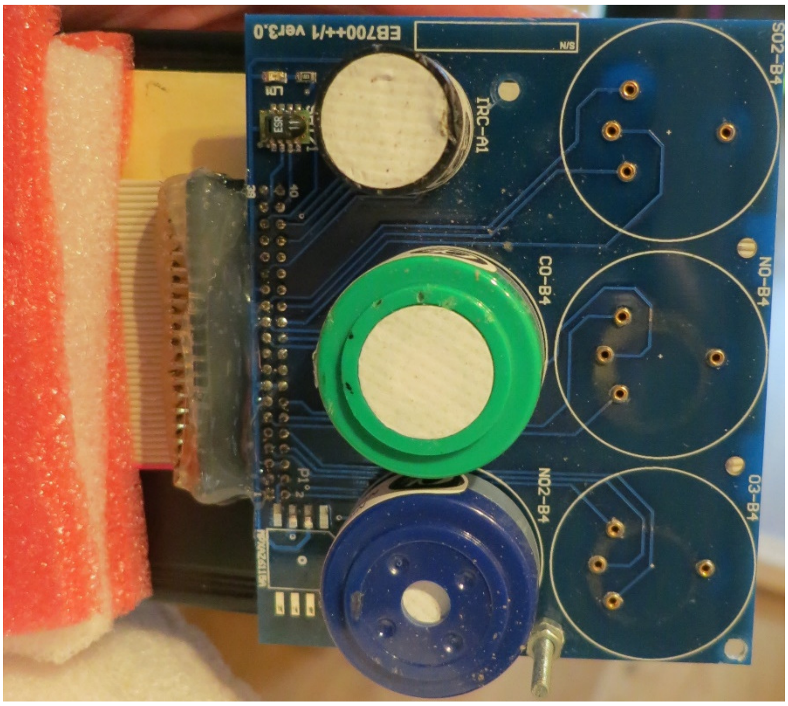
The data from the mobile monitoring campaign will thus be compared to the data obtained through the fixed-station monitoring to assess the differences and the validity of the data obtained from the campaign.

6.3 Pre Campaign trials

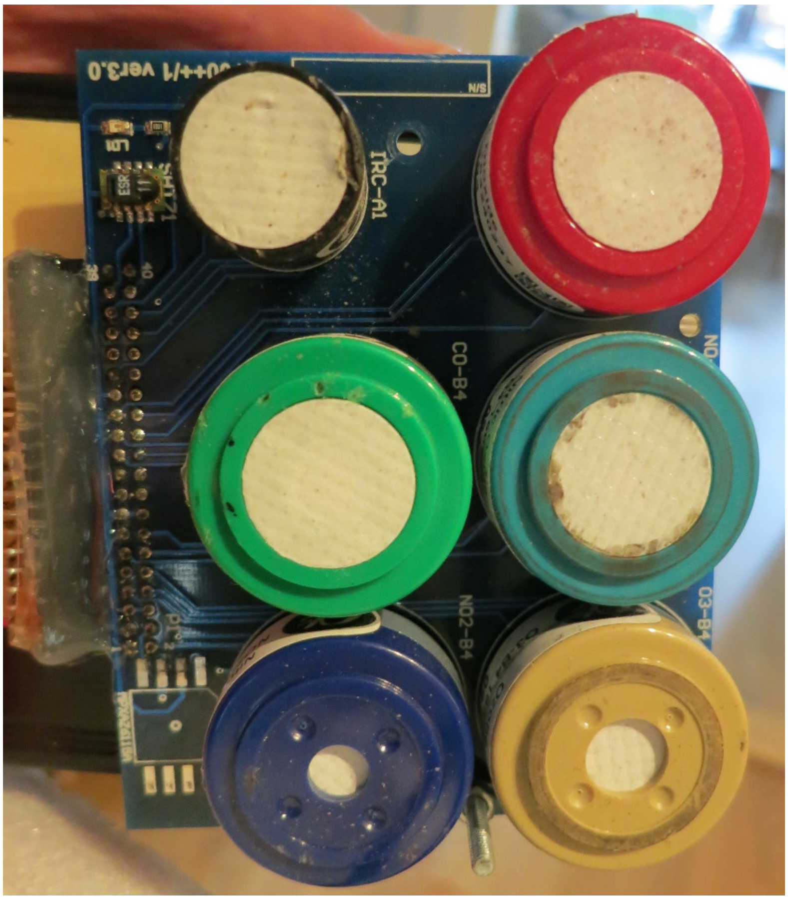
The monitoring equipment were tested in pre campaign trials in the autumn of 2014. During these trials, several issues with the platform became apparent.

The first issue discovered was that the platform provided data with a high number of no readings. Because of this discovery, the platform container were opened to inspect the status of the platform and the sensors visually. Several of the sensors had actually disconnected from the platform, and were spread around inside the container. The cause of this is unclear, but most probably due to vibrations from variations and holes in the pavement or other sources. The platform and the sensors would benefit from more padding inside the container to lower the vibrations obtained in the urban environment.

The sensors that had disconnected were reconnected, but new trials revealed that they had been damaged, and did not operate properly. The platform got new sensors mounted, but again, the readings obtained were in such a state that they could not be used. This meant that both the platform and the individual sensors had to be tested and recalibrated before it could be used in campaigns. Since this was the only bike available from the project, the mobile monitoring campaign had to be put on hold until the platform was ready again.



Picture 3- Sensors disconnected



Picture 4 - Sensors reconnected

Testing and recalibration of the second set of sensors was done in a NILU operated calibration lab during the winter of 2014/2015. The sensors were evaluated, and declared reasonably well performing.

A local testing campaign of new sensors were done near the NILU main offices at Kjeller, and the readings were verified to be in order. Additionally, a new Styrofoam protection casing for the platform were installed between the platform and the casing in order to minimize the impact of vibrations on the platform. As of this, the sensors and the platform were deemed ready for the mobile monitoring campaign without further pre-adjustments.

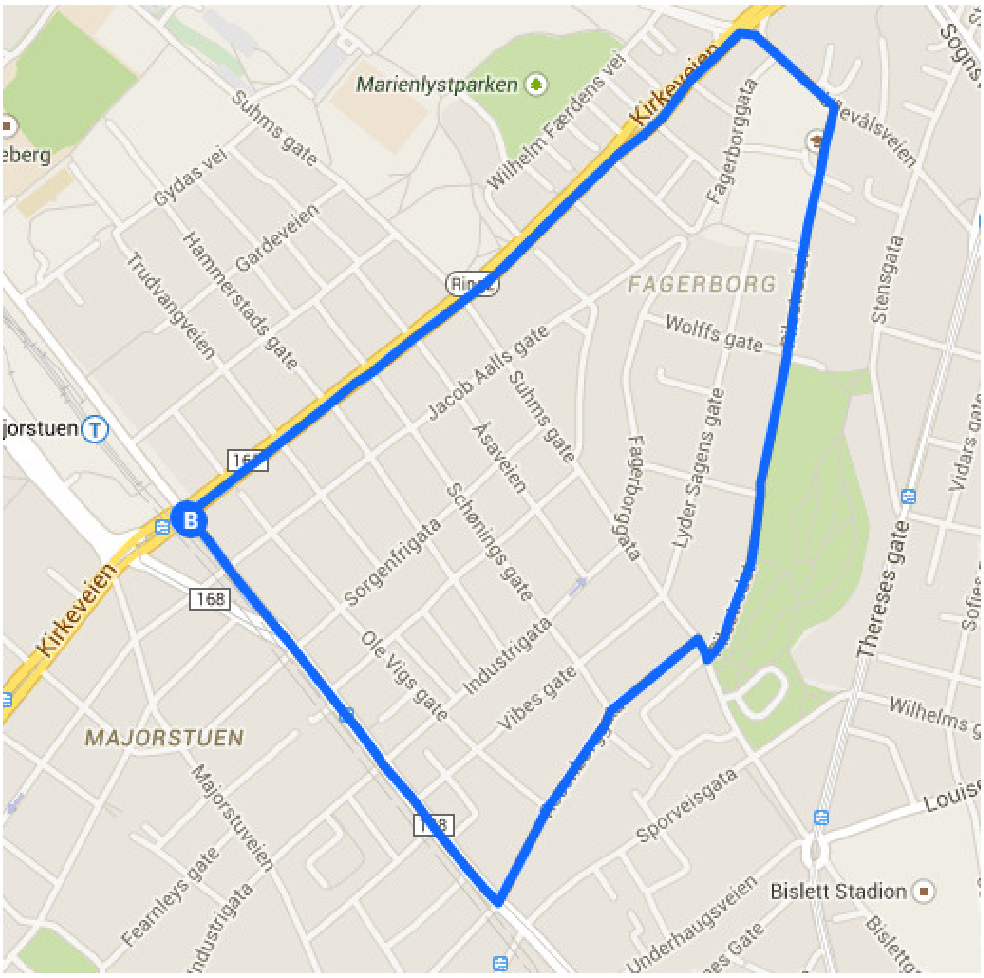
6.4 Mobile campaign in Oslo

The mobile monitoring campaign use the newly recalibrated platform, and follow two predefined routes at Majorstua. The first route is circular, and go along the streets of Kirkeveien, Pilestredet, Rosenborggata and Bogstadveien. The second route go inside the first route, along the parallel streets of the Fagerborg residential area at Majorstua. The platform monitor the levels of pollutants with a temporal resolution of 1 minute.

6.5 Kirkeveien – Pilestredet – Bogstadveien route

The first route is set up to address the spatial and temporal variations on a route combining densely trafficked roads with low trafficked roads. The route covers Kirkeveien from the major intersection at Majorstuen to the intersection with Blindernveien. At his intersection the route changes to the Ullevålsveien for 150 meters, then going down Pilestredet. It also go along the side of the Stensparken, a small park residing on a hill. At the intersection with Suhms street, the route changes to the Rosenborggata, a one-way residential street going to Bogstadveien. The route goes up Bogstadveien all the way to the Valkyriegata, a short road stub connecting to the Majorstuen intersection. This campaign is about 3 km in total length.

In the rest of this thesis, the first route will be referenced as the Kirkeveien route, and the second will be referenced as the Fagerborg route. The four streets selected for this route has varying traffic intensity and topology.



Picture 5 - Kirkeveien route - obtained from Google Maps

Kirkeveien

Kirkeveien is a major arterial road in Oslo, and is a part of the Ring 2 road system. It is trafficked by large numbers of cars and trucks, around 20500[37] every day, and is a major road for the bus network. It has two lanes each way for the traffic, bicycle lanes on the shoulder of the road, and a separated pedestrian sidewalk. In addition, it has several bus

[37] http://www.vegvesen.no/_attachment/234512/binary/436673?fast_title=Alleer+og+trerekker+i+Oslo.pdf

stops, where buses idle for some time while they drop off and pick up passengers. Medium rise residential and commercial buildings surround most of the road. The part this route covers is about 940 meters.

Pilestredet

The part of Pilestredet used in the campaign is a quiet street with sidewalks at both sides, and with cars parked at the shoulders. At the start of the street there is a high school, and following the road there is a road block to prevent traffic. At the end, the road follows the Stensparken, a small park residing on a hill. The part of Pilestredet used in this route has mostly low-rise residential buildings, and with little traffic except from the residents.

Rosenborggata

Rosenborggata is a part of the Fagerborg residential area, and is included in the second route as well. Rosenborggata is a one-way trafficked street connecting Bogstadveien to Pilestredet. The street has a slight elevation, which declines again when it meets Bogstadveien. It is part of the Fagerborg area, and has medium rise apartment buildings, some with service providers on the first floor.

Bogstadveien

Bogstadveien, one of the most popular shopping areas of Oslo, has high traffic, tramlines, and many pedestrians and cyclists, and a daily traffic count of around 11300[38]. It has also a high number of traffic lights and pedestrian road crossings. Every building has shops or businesses on the ground floor, and is a popular shopping area in Oslo. Recently the street was upgraded with wider sidewalks, to accommodate for more pedestrians.

Below is a summary of the different characteristics for the streets selected for the campaign. It describes the traffic conditions, city landscape, public transport and other special characteristics.

[38] http://www.vegvesen.no/_attachment/234512/binary/436673?fast_title=Alleer+og+trerekker+i+Oslo.pdf

The streets of the Fagerborg route are highly homogeneous to Rosenborggata in length, traffic density and building mass. As such, Rosenborggata describe all the streets for that route.

	Kirkeveien	Pilestredet	Rosenborggata	Bogstadveien
Traffic density	Very high	Very low	Low	High
Car Lanes	4	2	2	2
Cycle lanes	2	0	0	0
Sidewalks	Both sides	Both sides	Both sides	Both sides
Length monitored	940 meter	720 meter	435 meter	630 meter
Bus routes	20, 28	-	-	-
Tram routes	-	-	-	11, 12, 19
Other characteristics	Part of Ring 2	Run along park	One way traffic	High number of pedestrians

Table 4 - Street characteristics for the routes

Below is pictures from the streets visualizing the variations in traffic and building structure.

Picture 6 - Kirkeveien - obtained from Google Maps

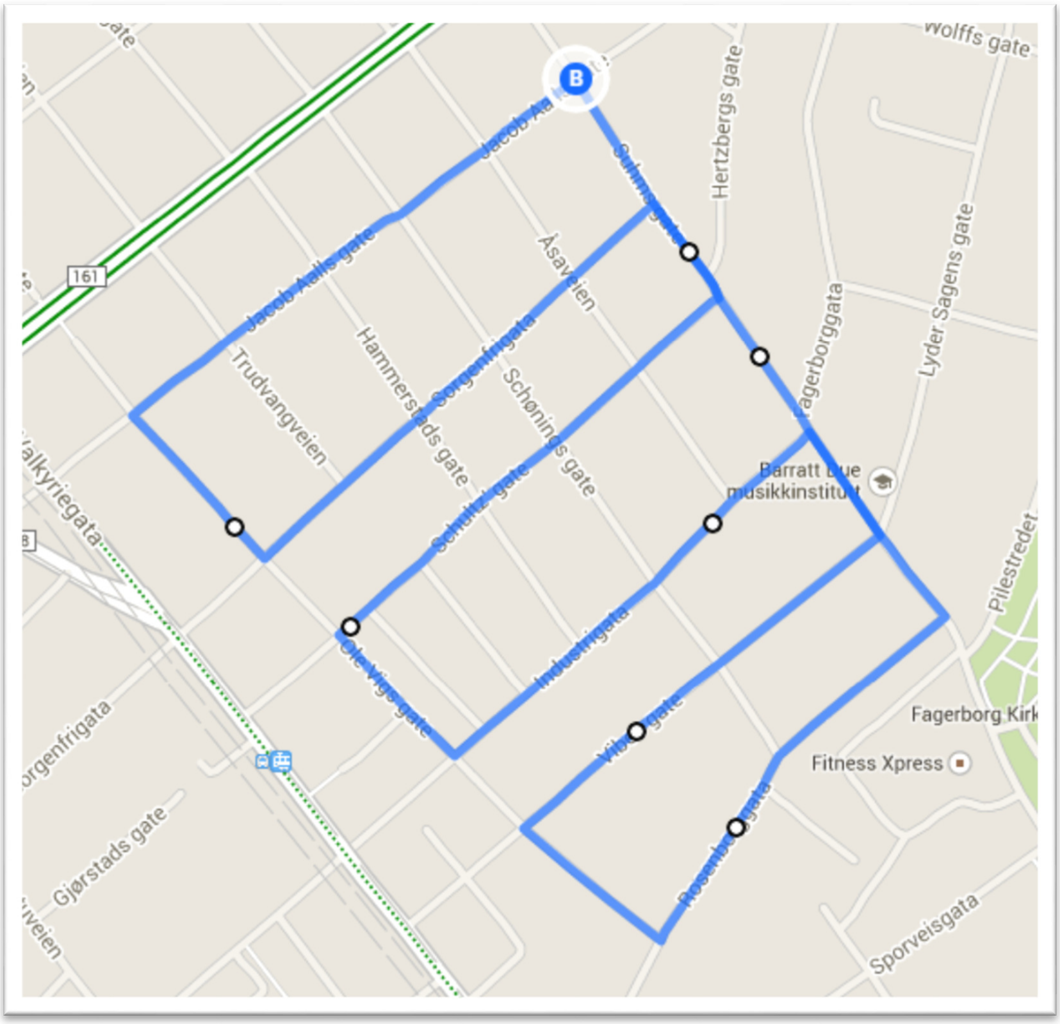
Picture 7 - Pilestredet - obtained from Google Maps

Picture 8 - Bogstadveien - obtained from Google Maps

6.6 Fagerborg route

The second route takes take place inside the first route, and provide a complete neighborhood saturation. The Fagerborg route covers the residential area of Fagerborg at Majorstua. The area is highly homogeneous residential area. The streets has two lanes for bidirectional traffic and pedestrian sidewalks, but without a separate bicycle lane. It has no bus stops or public transport, and has low utilization by others than its local users. Mostly medium to low rise apartment buildings surround the streets.

The route covers all the parallel streets of the area, starting with Jacob Alls gate. Then it goes through Sorgenfrigata, Schultz gate, Industrigata, Vibes gate and Rosenborggata, before it follows Suhms gate up to Jacob Alls gate again. Ole Vigs gate is used as a crossing street on left part of the route.



Picture 9 - Fagerborg route - obtained from Google Maps

Picture 10 - Industrigata - obtained from Google Maps

Picture 11 - Sorgenfrigata - obtained from Google Maps

6.7 Campaign execution

The mobile monitoring campaign was undertaken in the spring of 2015, between 08.04 and 20.04, with a total of 7 days of measuring. Each day the measurements was done for three hours between 13 and 16. These hours were selected to be able to get the variation in traffic density and conditions ranging from normal mid-day traffic to rush hour traffic. Additionally, both normal working days and weekend days were included in the study to assess the variations in the weekends to the variations found during normal working days.

The weather conditions during the campaign were mostly sunny and with low wind. The sunny weather resulted in some of the warmest days of the whole spring period.

In order to get the most data possible for all the hours of the campaign, the runs were repeated two times for each route every hour. On average, each run took about 15 minutes, with small variations where the bicycle had to adjust the speed to the surroundings accordingly due to the urban nature of the campaign.

6.8 Mobile measurement data analysis

After the campaign was finished, the data obtained needed to be analyzed to assess the possible temporal and spatial variations, and validate the readings obtained from the platform to the data obtained from the fixed-site monitoring station for the same period.

6.8.1 Data preparation

The analysis of the data was done with Microsoft Excel, R Studio version 0.99.441 as the R IDE and R 3.1.3 as runtime. The computer used is a PC running Microsoft Windows 8.1.

The initial exploration of the data from the platform was done in Microsoft Excel. The later statistical analysis and plotting was done in R Studio, an IDE for developing scripts and applications for the R language.

R[39] is a programming language and software environment used in statistical computing, developed as a free alternative to the S language. The use of R has risen recent years, mostly due to the recent availability of large datasets, along with increased interest in statistics and data mining of large datasets.

6.8.2 Data extraction

For the data extraction, all the data for the days of the campaign was downloaded into one file. The Dunavnet[40] site enables users to log in to get the data for their platforms. It shows the last recorded status for the measuring platform, and lets the users download a file with all the data for a given period. The site shows only status of the last measurements, and has no capabilities to analyze the data in itself.

To be able to split the dataset into the different road segments, the location coordinates had to be resolved into streets. This was done manually by copying the coordinates into Google Maps and writing the street into the datasheet. Location coordinates that were outside the campaign route were manually calculated to the nearest street.

The electric bike is continually measuring the air quality when it is turned on. This means that the values measured during travelling back and forth to the campaign routes had to be removed from the dataset. To be able to distinguish the values recorded on the separate routes and the record not on route, a column that stated the route of the record was added to the dataset.

Due to instability of the GPS sensor, some record were missing the location coordinates. These records also had to be marked in the dataset before the analysis. This was done by adding a column to the set that stated possible errors such as GPS failure.

Below is a table describing the number of initial reading obtained, and the filtering steps needed before the records for the two routes were ready for analysis.

---

39 http://cran.r-project.org/doc/FAQ/R-FAQ.html#What-is-R_003f

[40] citisense.dunavnet.eu/index.php

	Removed	Total left
Initial number of records		1089
Removed records without GPS	271	818
Removed off route record	125	693
Total number of records for Kirkeveien route		358
Total number of records for Fagerborg route		335

Table 5 - Record filtering process

Breaking down the multidimensional dataset

To prepare the dataset for analysis in R, the dataset was cleaned and divided into a series of subsets. Records without valid location coordinates, or did not belong to either of the routes were filtered out, creating a complete set of valid records. The complete set is then split into subsets for the two separate routes, and the routes were split into datasets for each street. This made it possible to analyze the spatial variations such as street-to-street variations.

To analyze the temporal variations along with the spatial, the street subsets were split into hour-oriented subsets, and into date-oriented subsets. For analysis of the temporal variation in hours for the different streets, the road segment subsets had to be divided into more subsets. The new subsets contained only the data of a given road segment for the different hour intervals of each route. The same was done to get the temporal variation in days, but the split was done for the different dates instead of the hour. The initial dataset and the datasets for the Kirkeveien and Fagerborg route is available in Appendix D.

6.8.3 Quantitative data analysis

Temporal-Spatial variance

The datasets derived from the sub setting holds the different dimensions that the analysis can be based upon. To analyze the variance in time for the different streets, the mean for each pollutant at the different hours for each of the road segments were calculated. These mean values are compared to assess the variations. The street-to-street variations in percent were calculated. This was done for all the different pollutants. In addition, the standard deviation was calculated, and the relative standard deviation in percent was added to the dataset.

Street oriented spatial variations

To assess the variation between the streets, the mean recorded at Kirkeveien at 14 was compared to all the other street means. This enables one to see how the pollution levels of the highest trafficked street compare to the other streets of the campaign

Hourly temporal variations

To assess the variations between the different hours for the different streets, the mean at 14.00 was compared to the means at 13.00 and 15.00.

Daily temporal variations

To assess the day-to-day variations, the daily mean of the pollutants was compared to the other days.

6.8.4 Comparative analysis with reference location data

In order to compare the values of the platform to the data from the fixed-site reference station in Kirkeveien, the PPM values from the platform needed to be converted to ug/m3.

The reference monitoring station measure the air pollution in micrograms per cubic meters (ug/m3), while the bike platform uses the parts per million (PPM) scale. In order to compare the measurements of the bike platform to those of the fixed station, the PPM values obtained from the platform need conversion to ug/m3.

The measurements in PPM are first converted to PPB (parts per billion), and then converted to ug/m3 with specific conversion algorithms for the different pollutants. NILU provided a script for the conversion, and the algorithm use the various pollutants molecule mass and the temperature at the measurement time.

The conversion was done with the values for NO_2 collected at Kirkeveien at the different day intervals. Means were calculated for each campaign day, and the values were then compared to the corresponding values collected at the fixed station.

Date and time	Means from bike platform	Means from the fixed station
08.04.15 13.00-16.00	4947,37	43,66
09.04.15 13.00-16.00	5793,26	59,94
16.04.15 13.00-16.00	5325,41	29,43
18.04.15 13.00-16.00	5125,63	31,35
19.04.15 13.00-16.00	5166,16	16,51
20.04.15 13.00-16.00	5209,44	78,31

Table 6 - Comparison table of means in ug/m3

The conversion revealed that the values obtained from the platform was unrealistically higher than the reference values from the fixed-site station. Further investigations revealed that the value platform states as being PPM needs recalculation with calibration data obtained from calibration test.

NILU performed the calibration tests in a professional gas-chamber lab, and after obtaining the calibration data from the lab, the values from the campaign was recalculated for new analyses.

The initial spatial and temporal analyses were now of little value, since they were based on data that needed recalculations. However, new analyses on the calibrated data similar to the ones already produced could possibly still produce insights on the temporal and spatial variations.

6.8.5 Relative data analysis with lab corrected data

The new calibrated values revealed that the data obtained in the campaign actually had so low quality that it had no value. For the NO_2 and O_3 sensors, there was a large amount of negative readings.

A comparison of the high trafficked street of Kirkeveien to the low trafficked street of Pilestredet was done to investigate how the NO_2 and O_3 sensor values varied on a street-to-street comparison. There were little to no observable street variations between the streets, even with the largest difference in traffic. The means of the streets were also calculated, and the mean for the very low trafficked street Pilestredet were higher than the mean from the high trafficked street Kirkeveien.

The spatially oriented plot below show the high variations in the measured values for the different streets. As previously mentioned, it show that the mean measured at Pilestredet is above the mean measures at Kirkeveien. The red line shows the null, and all the points below that line represent the negative readings.

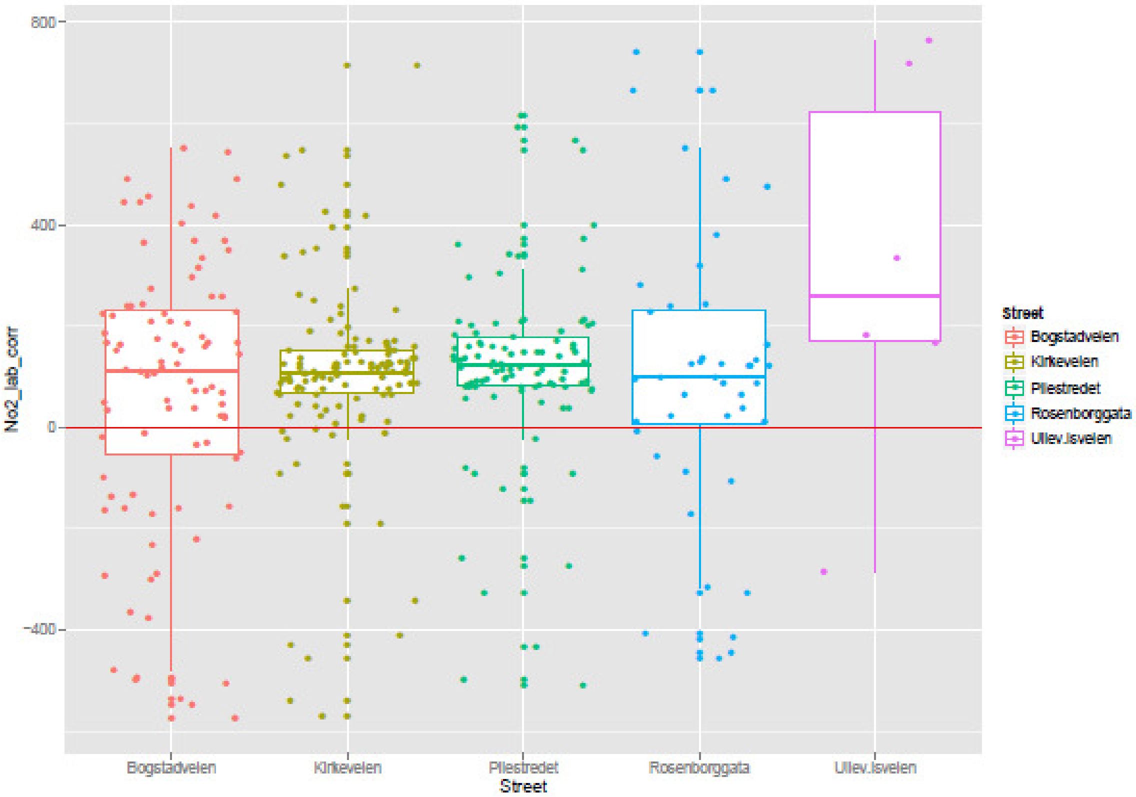


Figure 3 – Boxplot Streets/NO_2 - Kirkeveien route

The temporal oriented plot below show the O_3 levels obtained at the hours for the Fagerborg route. Here are the means decreasing, contradictory to the diurnal traffic patterns. There are also a high number of negative readings, as can be seen by all the points below the red line.

Figure 4 - Boxplot Hours/O_3 Fagerborg route

In Appendix A and B is a set of boxplots describing the dimensions hour, day and street for the NO_2 and O_3 levels, both for the Kirkeveien and the Fagerborg route.

These analyses shows that the data obtained in the campaign is invalid, and thus unable to show the spatial and temporal variances of air pollution. The reason for the failure in the readings need further examination, as there is a multitude of possible reasons. There could be errors from the sensors themselves, the platform can have errors, or the mobile nature of bicycle-based air quality monitoring might have implications on the sensory readings.

NILU performed a post-campaign analysis of the sensors, and by conferencing with the providers of the DunavNet platform, it appears as the sensors were giving faulty readings due to aging of the sensors. The lab calibration data is available in Appendix C.

6.8.6 Lessons learned

Studies of using low-cost sensors for mobile monitoring of air pollutants is to this day low in number, and further evaluation of different sensors for air quality monitoring and their performance during mobile use are needed. The quality of the data provided needs to be at a trustworthy level both for private and public use.

While the anecdotal nature of the personal experiences of the campaign do not have scientific value, they can especially in pioneering campaigns like this, provide insight into areas that would benefit from further exploration.

One of the experiences from using the platform in the campaign was that it is hard to know of the platform was working properly. There was a light indicator for power on the connecter board of the sensors, but to be able to see it, one has to get off the bike and peek in under the air vent. The light was in addition blue, making it hard to spot in bright daylight.

There was no other indicators for the status of the platform. The GPS signal was lost several times during the campaign, and one whole day of monitoring was lost due to no GPS data. This can be very frustrating and reduce the commitment to the campaign, especially for campaigns relying on community involvement. Light indicators for power, GPS, mobile data and platform status can help to know the real-time status of the equipment, so the equipment can be adjusted and repaired in case of failure. Additionally, the platform could benefit from tighter integration of the GPS and GPRS modules, as they now are separate units that can disconnect without notification to the user.

The notion of using bicycles as the transport mode to monitor the variations of urban air quality has its own inhabitant limitations. The need for power for the monitoring equipment can be solved by either separate batteries for the equipment, or using the power available on electric bikes. This sets constraints on the duration of the campaigns as the batteries need reloading after each use.

Repeated runs of a predefined routes is the recommend approach from similar studies, and a high number of repetitions is needed to obtain representative results (Van den Bossche et al. 2014). This repetition of the same routes day after day makes the monitoring both a tedious and tiresome experience. Additionally, this can be an issue for campaigns relying on community commitment, as the community might not adhere to the scientific constraints of the campaigns.

The weather and season during the campaign also have impact on the choices of the timeframe of the campaign. Although this campaign was done in nice and sunny springtime weather with good conditions for cycling, it can be possibly hazardous to be monitoring in high trafficked winter streets with snow and icy conditions. Therefore are bicycle campaigns limited by the weather and season, and data from winter months will be scarce.

6.9 Conclusions from the campaign

The campaign revealed that the platform, at its current state, need further studies and improvements to be able to monitor the air pollution correctly, and be used in mobile monitoring campaigns.

6.9.1 Data validity

Sensor recalibration

As seen from the campaign, the solid-state gas sensors are susceptible to the harsh environment of urban mobile monitoring. In order to function properly and deliver quality data the sensors needs recalibration at a given interval. A study of similar low-cost sensors from the OpenSense project (Hasenfratz et al. 2012) states that recalibration due to sensor aging is needed every month or even every week to ensure the validity of the sensors. The traditional way of recalibration requires a gas-chamber calibration lab, normally only found in labs highly specialized in air monitoring, and require trained staff for operation. This can be a challenge, due to the low availability of such labs, and the high cost and time consumption.

Hasenfratz et al. (2012) propose a solution to the challenge of calibrating the sensors through on-the-fly calibration algorithms. By utilizing the GPS module information and the sensors encounters with other low-cost sensors or fixed-site monitoring station, the sensors can use the rendezvous to compare each other's readings to improve their calibration and increase their accuracy.

Reference data

In order to assess the quality of the data obtained from mobile monitoring campaigns, reference data must be obtained from a fixed-site monitoring station. The quality of the data from the fixed-site stations has a high degree of trustworthiness, as they are robust high quality monitoring equipment tested over several years, and are continuously maintained and calibrated. The availability of the reference data is crucial to monitoring campaigns, and it is used in this thesis's campaign as well as the other mobile monitoring campaigns studied. As the fixed-site monitoring stations is of low number and set at strategic monitoring sites, this limit the possibilities for route selection in mobile campaigns.

Monitoring methods

The validity and trustworthiness of the data provided is crucial for mobile monitoring campaigns, due to the novel nature of the campaigns. There are skepticism in the environmental monitoring community to the data provided, on topic such as the quality of the data, but also on the aims of such studies (Aoki et al. 2009). Therefore, the data obtained must be of sufficient quality for scientific use, and the methods for campaign execution must be followed rigorously.

6.9.2 Data resolution

The resolution on the data obtained from the platform from the campaign is about every minute. The other mobile monitoring studies examined has the resolution set to between 1 second and 30 seconds. While this is both dependent on the equipment used and the desired data, the resolution is of great importance for mobile monitoring and air quality mapping. The speed to the transport vehicle along with the data gathering resolution effectively regulate the amount of data obtainable from the campaigns. For bicycle transport, the speed is normally around 15 km/h. By converting the speed to meters per second, the speed is 4.17 meters per second. For a resolution of 1 minute, the distance traveled between the data points gathered is roughly 250 meters. This low spatial resolution results in considerably larger number of runs needed to gather the data points necessary to measure the spatial variations of air pollution than for equipment with higher resolutions such as 1 second.

In the mobile monitoring best practice study performed by Van den Bossche et al. (2014), they have found large variations in air pollution levels over short distances. The study states that to be able to map the quality of the urban air and identify the hotspots or peak areas, a resolution of 20-50 meters is required.

6.9.3 Weather dependent monitoring

Mobile monitoring campaigns that employ bicycles as the transport vehicle rely on weather conditions that allows for cycling. While this is not an issue in many countries, the cold climate in Norway effectively stops such campaigns for several months of the year. During the winter and the months before and after, the low temperatures along with snow and ice on the roads, make bicycling in the urban environment a cold experience at best and hazardous at worst. These months are therefore hard to monitor with bicycles, and other monitoring approaches are necessary to gather data of the air pollution levels during these months.

6.9.4 Equipment status

In order to improve the user friendliness and confidence in the platform, it needs to improve the communication of the status to the user. Simple status LEDs can inform the user that the platform is actually monitoring, and the status of the GPS and GPRS connection.

A mobile app could extend the information of the platform and display the measurements in real time. This app could also alert the user of issues with the equipment.

7 Approach for an Urban Air Quality Recommender System

This chapter present an approach for an Urban Air Quality Recommender System. Building on literature findings and campaign experiences, it will present an approach for a context-aware knowledge based recommender system based on real-time air quality data.

As found in the literature, the advent of low-cost sensors for air pollution enable a new air quality monitoring paradigm, providing insight into the actual air pollution variability in the urban environment. However, this insight needs to be conveyed to the public in ways that are personally relevant. Through an air quality recommender system the public will be able to utilize the information, and actually make informed decisions that lower their exposure.

Due to the low quality of the data from the campaign, this thesis will not be able to use the campaign data to test any models. This opens up for a more visionary approach, and this chapter will describe an envisioned context-aware recommender system on a high-level, and will include some services not yet available.

Time

Time is a crucial element of the decision-making process for most people. Time set constraints on the everyday events, such as getting to work or other appointments. There is a Norwegian phrase dubbed "Tidsfella" which refers to the feeling of not having enough hours in the day to do all the planned tasks.

While being a diminishing resource in every person's life, taking time to go a longer but healthier route could actually lower the personal exposure and improve the physical condition, thus giving back a generally improved health status and a prolonged life. In addition to the time and health aspect, the healthier route is often situated in more beautiful surroundings, as parks and recreational spaces, enriching the traveling experience.

The system aims to give route directions along with the air quality information for that route. Additionally, the system ask the user if they have time available to go a healthier route, and based on the amount of time stated by the user, the system will calculate less polluted route alternatives within the given timeframe.

7.1 Functional system description

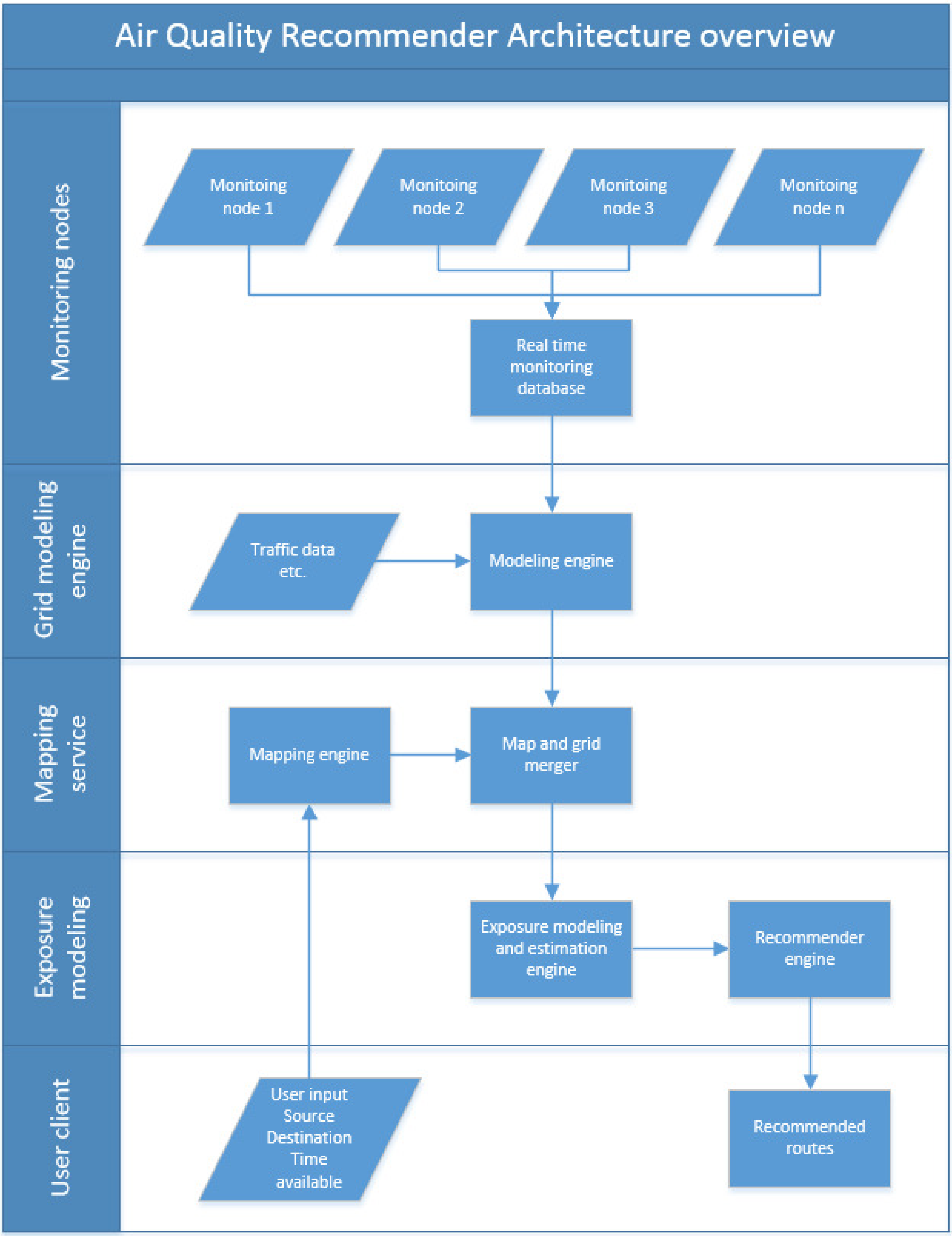
This recommender system envisioned will let cyclists make an informed choice of travel route based on high-resolution air pollution data. Based on input from the user on their desired destination and their available time, the system will use the personal contextual data along with modeled air quality data from real-time monitoring and historical data to create an air quality map of the route. The system calculates the fastest route to the destination, and an alternative route less affected by air pollution within the given timeframe.

7.2 Functional architecture

The functional architecture of the recommender system encompasses several models and algorithms. First, the selection of the type of recommender model is presented, along with constraints for the recommendations. Secondly, the algorithm for adding the contextual data is presented, and then a description of the data required by the model.

The system will have a classic server-client architecture, where the servers do the modeling and analysis, and the user client is a smart phone application or web service.

Below is an overview of the system architecture, including the real-time monitoring and modeling, the mapping engine and grid model merger, the exposure modeling, and the recommendation engine.



7.2.1 Recommender model

The envisioned system will be based on a knowledge based recommender system model. This choice is due to the fact that the recommendations are primarily based on the knowledge of the levels of air pollution. The most important input will be the real-time monitoring data, along with modeled data for the areas not covered by mobile monitoring.

The knowledge-based model will also employ a constraint based recommendation technique, which uses the input of the user to provide suitable recommendations. The constraint given by the user is the choice of route and timeframe available. Based on these constraints and the system knowledge, the model will create a set of recommended routes for the users.

7.2.2 High-resolution air pollution grids and maps

Real-time data are gathered and stored in a central database containing all the monitoring data, both real-time and historical. The real-time data can then be used to create a high-resolution gird for the monitored areas. As seen by Van den Bossche et al. (2014) pollution variations is detectable down to a resolution of 20 to 50 meters, thus can the grid be of similar sizes, depending on the resolution of the data provided.

Modeling engine

To model data for the grid cells without or with too few measured data points, one can employ the modeling technique presented by Hasenfratz et al. (2014). This technique uses land use regression with additional traffic data to model the air quality for the grid cells without, or with too few measured data points. Additionally, the model is also able to make use of historical data that has many of the same characteristics, such as similar meteorology factors, to improve the model and the grid quality.

Map and grid merge

This pollution grid can be merged with the desired route map as an overlay. This combined map will then provide a basis for estimation of the air pollution exposure.

Data fallback

Traditionally modeled data can be used as a fallback if the real-time monitoring fails to provide the data foundation necessary. Still, the resolution is so low that the system only will be able to provide general recommendations for large areas, like the ones available through the public information site luftkvalitet.no, and are thus not able to provide the fine granularity needed by the recommendation system.

7.2.3 Domain knowledge

The domain knowledge acquired in chapter three can be used to select the relevant data for this system. The knowledge describe the potentially dangerous pollutants in the urban environment, and their implications on human health. It also describe some of the various factors that affect the spatial variations in urban environments. The domain knowledge gathered need to be adjusted to integrate new knowledge, as the field of urban air quality and health effects are under research, and new knowledge are becoming available.

7.2.4 User input

The system need to gather various information from the user to be able to start the recommendation process. This information must be obtained both explicitly and implicitly, both before and after the recommendation is done and the user choice is taken.

Start location

This will be obtained implicitly by obtaining the location coordinates from the mobile phone or explicitly by letting the user provide a start address.

Destination

Since the user types in the destination for the journey, the destination data is obtained explicitly from user input.

Timeframe available

The timeframe available for the users is entered explicitly, either in available minutes or as a percentage increase in travel duration.

User choices

The choice taken by the user on whether to choose the fastest or the healthiest route is interesting both for assessing the performance of the recommendations, but also the patterns of use can deliver insight for researchers, governmental agencies and the developers of such systems.

7.2.5 Exposure estimation

To estimate the personal exposure for the proposed routes, the system can use the model proposed by Hasenfratz et al. (2014), described in section 4.5 . The model relies on a grid model of the air pollution levels, and calculates the exposure based on the levels of air pollution measured and the time spent in the grid cell.

The total exposure for each route can then be calculated by summarizing the grid cells for the desired route. This rather simple model will be useful for a recommender system, as the computational resources needed are generally much lower than with traditional modeling techniques.

7.2.6 Route visualization

The final recommendations can then be visualized to the user in the form of a map with the suggested routes and their pollution levels. As the different levels of pollution in the form of ug/m3 or PPM can be difficult to assess for a common user without insight into air pollution levels, the system has to communicate this in a simplified way.

Air Quality Index color codes

The AQI has been available for several years, and is known as a simplified expression of the measured levels of air pollution from the fixed-site monitoring stations. The color codes from the index can be used in the route recommender system, to mark the routes with the color code applicable for the different levels. This will make the users aware of the pollution on the routes, and help them make a healthier choice.

7.3 Possible extensions

7.3.1 Improved exposure estimation

Extending the recommender system with additional and optional user input can adjust the recommendation for a better fit to the user preferences. The mode of transport influence the personal exposure to air pollution due to the factors described in chapter 3.5.

Mode of transport

The mode of transport has implications on the personal exposure, as the heart and respiration rate changes due to the user activities, and the circulation of air varies greatly from inside vehicles to the outdoors.

This recommender system, focusing on recommending healthier alternative routes for cyclists, has the possibility to extend to other modes of transport. However, this requires additional data from the other transport modes, as Peters et al. (2014 p.41) states that *"cyclist exposure maps might not be representative for other transport modes. Several studies (e.g. De Nazelle et al., 2012; Dons et al., 2013) have demonstrated the impact of transport mode on exposure, but the potential of extrapolating mobile measurements from one transport modus to another has, as far as we know, not been studied in sufficient detail."*

Health status profile

Information on the health status of the user has the possibility improve the recommendations, and to be more specific in their personalization. The recommended exposure thresholds for persons affected by respiratory and cardiac diseases are lower than for the public, as are the health effects more severe. Although this information will be of great benefit for making recommendations, it also have issues regarding privacy and security of personal health information. To address this problem, and still get information the user's health status, a point-based scaling system can be applied.

By adding a health status profile system in the user client, the system can use this data to adjust the AQI color codes to match their health status profile, after the recommendations are received from the system. The system, based on the willingness of the user to complete such a profile, can make a profile that matches the current health status of the user. After the profile is complete, a personal health rating is provided to the user. This rating can be of 1 to 5 points, depending on the severity of their health challenges. This simplified point ranking can then be applied to the recommendations to adjust the AQI color codes for improved personalization of the recommendations. Since the information is on the user phone, and the adjustment are done after the recommendations is given using a point based scale, the health data will not be obtained by the recommender system, protecting the privacy of such information.

7.3.2 Building user profiles

The system collect data on the user selections and the overall use of the recommender system. This data collection is displayed to the user as the profile grows, and might visualize various graphs and statistics on their selection and both their total exposure, the exposure reduction, and personal goal achievements.

By analyzing the user data one can create statistics of the use and performance of the system, and analyze the user choices. Such data can also be of interest to the research community, governmental agencies and the system developers.

7.3.3 Opening the data

Providing access to the mobile monitoring data for external interests provides opportunities to integrate the data into applications and use cases not yet explored. The Future Work section of chapter 8 will present some of these possible use cases.

7.4 Challenges

To make such a system a reality, there is several challenges that has to be addressed. There are challenges in the form of the air pollution data, testing of the models and their integrations and user acceptance and use of the system.

7.4.1 Monitoring data foundation

The most important data for the system is the monitoring data. The data has to be reliant and of acceptable quality, and have the spatial resolution necessary to portray the actual levels of air pollution down to street levels. As the results from the campaign show, there is still work to be done to make such monitoring a reality.

7.4.2 Model testing

The models proposed need further testing with real data sets to assess their performance and integrations. Some testing has been done in the study by Hasenfratz et al. (2014), but this has to be extended with other datasets from mobile monitoring.

7.4.3 User acceptance

To know if the users actually will use such a system, a user evaluation of the system should be performed. This can reveal obstacles, challenges and possibilities in the use not foreseen in the development.

8 Conclusion & Future work

This chapter will first examine the conclusions of the campaign questions and challenges, and the personal experiences from the campaign. Then it will investigate possibilities for future work in mobile monitoring and recommender systems based on mobile monitoring data.

8.1 Conclusion

Due to the lack of valid data from the campaign, the research on the spatial and temporal distribution with low-cost sensors will remain unexplored. Therefore, the evaluation of the personal experiences and lessons learned will form the basis for most of the conclusions. These experiences will address the challenges faced from the equipment, in regards of sensor stability, data resolution, equipment protection and status information.

8.1.1 Spatial and temporal air quality variations

As previously mentioned, the analysis of the spatial and temporal variations could not provide accurate data and thus will not be usable to any analysis. As found in the literature, there has been discovered variations in air pollution down to the intra-street level. This make mobile monitoring projects still viable, and with great potential for high-resolution air pollution data and maps. Still, one of the main issues is to get the low-cost sensors to perform adequately in the urban environment, and remain accurate over time.

8.1.2 Evaluations on lessons learned

By using the sensory platform from the Citi-Sense MOB project, this project has pushed the envelope for obtaining mobile monitoring data with low-cost sensors. However, as in many other pioneering projects, this has not been such a success as initially envisioned. Therefore, the personal experiences and the lessons learned from the campaign will be of higher importance, as they can provide insight into a domain that has not yet been studied in great extent.

8.1.3 Improvements

There are several possible improvements, both to the sensors and platform and to mobile monitoring campaigns in general and community sensing in particular.

Sensor recalibration

It is clear from the campaign that the sensors and the platform need further investigations for use in mobile monitoring. The issue with aging of the sensors and necessary recalibrations make the utilization of low-cost sensors for such environments challenging, and as discovered in the study by Hasenfratz et al. (2012), recalibrations is necessary on a monthly to weekly basis. The study propose an on-the-fly recalibration, using fixed-site measurements and the data from a network of sensors to recalibrate the sensors in real time. This can be a promising solution, but as the current number of operational sensory platforms remain low, this is not yet possible.

Sensor types

Low-cost high-resolution particle matter sensors, while being not utilized on the campaign platform, could be of high value to mobile monitoring as particle matter of different sizes are considered now as one of the main culprits for the negative health effects of air pollution. Particle matter is the focus for several of the campaigns studied, and while these studies employed high-cost equipment, low-cost alternatives could provide larger distribution of monitoring equipment, providing high-resolution data for these dangerous pollutants.

Data resolution

The temporal resolution that was obtainable from the platform used in the campaign was as low as one reading per minute. As discussed earlier, the other campaigns studied uses equipment that can provide a resolution of 1 to 30 second intervals. The resolution is of great importance to the spatial and temporal dimensions of the data in mobile monitoring campaigns, and with such a low resolution as provided by the platform, a very high number of repeated runs are necessary to obtain a useable data foundation.

Monitoring equipment

The monitoring equipment used in this campaign has had several breakdowns. First was the problem with sensor disconnection, most probably due to pavement vibrations. Due to this, the platform had new sensors installed, and calibration of these had to be done in a gas-chamber lab. At the same time new protection encapsulating the platform was added to the case to prevent such happenings in the future. The new sensors with the added protection did not disconnect, so it seems like this new protection solved this problem, but as the monitoring were done in a relative short interval, further examinations is necessary to verify the operational stability.

There were also issues with the connection of the cable for the GPS unit. A whole day of measurements was lost due to disconnection. Providing information to the user on the operational status of the equipment will therefore be of benefit, as equipment failure has a great impact on the dedication for mobile monitoring.

Recommender system

A recommender system like the one proposed in this thesis has the possibility to inform the public on the levels of air pollution exposure while travelling, and suggest other healthier routes. This can both increase the public knowledge of the air pollution in their surroundings, and decrease the personal exposure while travelling. However, such a system rely on high-resolution data from mobile monitoring to be able to calculate the exposure. As the data foundation is not yet available, such a recommender system will remain at the conceptual stage.

When the real-time air pollution data foundation is present, and given the amount of money calculated to smart city development this might not be far away, there will be ample opportunities to develop such systems.

Community sensing projects

Community sensing projects has the possibility to let the citizens take action and study their own environments. This sounds promising, but as discovered from the literature review, it is very important to follow the campaign methods strictly to attain the quality and the resolution needed from bicycle mobile monitoring campaigns. A high number of repetitions of the same routes is necessary, and getting the monitoring volunteers to follow the methods rigorously will require high dedication. Experience from the campaign execution tells that this can be a challenge, as cycling the same rounds over and over is monotone and unexciting, and can reduce the dedication to follow the campaign methods. Additionally, the intentions and motivations of the participants might be set on differentiating goals. This could lead to falsification of the data and misdirection of the monitoring, as stated in the study by Aoki et al. (2009)

Thus might the campaigns who utilizes monitors mounted on public transport following given routes or other similar public vehicles be a better alternative to provide high-resolution data on the levels of urban air pollution as seen in the Open Sense project (Hasenfratz et al. 2014). This way the scientific methods are easier to propose and follow, leading to more substantiated results.

8.2 Future work

While still being in its inception phase, mobile monitoring with low-cost sensors shows potential to be able to provide real-time high-resolution data on the variations of air pollution in the urban environment. Still, this thesis highlights several issues that need further research and testing in the urban environment

8.2.1 Recommendations

The recommender system envisioned in this thesis is merely on a conceptual stage, and need further development and testing to assess the real value and success for the recommendations.

8.2.2 Improved platforms

The campaign conducted for this thesis clearly show the reliability of the sensors needs further investigation and additional development to ensure the data validity and stability. Another point is that the resolution of the gathered data needs improvements to provide the granularity needed. Additional research and development on these topics will be of great benefit for mobile monitoring.

8.2.3 Improved modeling

This thesis has only found one study that has employ modeling to the data gathered to assess the pollution in areas not covered by the monitoring. The field of mobile monitoring will benefit from more campaigns and datasets to further investigate the accuracy of the model, and other models could be developed.

Exposure estimation on the mobile monitoring is also in need for further testing and verification as the model is rather simplistic, and might not be representative for the actual exposure.

8.2.4 Improved visualization

This thesis proposes a recommender system for visualizing the pollution levels for the users travels. As this is merely an example of a possible use, there is possibilities of a wide array of possible applications for such data. In addition, through opening up the data via API's, a multitude of possible uses opens up.

8.2.5 Possible extensions

Real-time data on the levels of air pollution can be beneficial and applicable to several other domains, such as quantified-self and personal assistant solutions. Additionally, mobile air monitoring could be extended by integration into smart city scenarios and solutions.

Quantified self

Due to the rise in the quantified-self domain, exposure information could be beneficial to various solutions such as Strava[41] and Endomondo[42], integrating the air pollution data into the existing solutions. This could extend the knowledge in these systems, and raise the awareness of the users on their levels of air pollution exposure.

[41] https://www.strava.com/
[42] https://www.endomondo.com/

Personal Assistants

Apple's Siri[43], Microsoft's Cortana[44], and other personal assistants are being developed to assist the users in finding relevant information, and perform various tasks automatically. Such systems could benefit from the availability of real-time data on the air pollution and their integration into other information sources such as the users calendar. Combining data like this can make the system assess the available time before the next appointment, and suggest a healthier travel mode or route.

Smart city infrastructure integration

Mobile monitoring of air quality could be integrated into various solutions in smart city infrastructures. Such infrastructures are able to monitor and control the several aspects of infrastructure, as traffic flow and speed. Information of the urban air quality could then be utilized to direct traffic away from highly polluted areas, by using rerouting or road pricing schemes.

Several cities around the world are testing various smart city solutions. In the city of Santander in Spain, an ongoing project[45] examines how to integrate air pollution monitoring into such solutions, and it will be exciting to see how these solutions develop in the future.

[43] https://www.apple.com/ios/siri/
[44] http://www.microsoft.com/en-us/mobile/campaign-cortana/
[45] http://www.smartsantander.eu/

Reference List

Abbey, D.E. et al., 1999. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *American Journal of Respiratory and Critical Care Medicine*, 159, pp.373–382.

Adomavicius, G. et al., 2005. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems*, 23(1), pp.103–145.

Adomavicius, G. & Tuzhilin, A., 2011. Context-Aware Recommender Systems. In F. Ricci et al., eds. *Recommender Systems Handbook*. Boston, MA: Springer US, pp. 217–253. Available at: http://www.springerlink.com/index/10.1007/978-0-387-85820-3.

Adomavicius, G. & Tuzhilin, A., 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), pp.734–749.

Allemann, D. & Raubal, M., 2013. Towards Health-Optimal Routing in Urban Areas. , (HEALTHGIS 13).

Aoki, P.M. et al., 2009. A Vehicle for Research: Using Street Sweepers to Explore the Landscape of Environmental Community Action. , p.10. Available at: http://arxiv.org/abs/0905.0200.

Bell, M.L., Davis, D.L. & Fletcher, T., 2003. A Retrospective Assessment of Mortality from the London Smog Episode of 1952: The Role of Influenza and Pollution. *Environmental Health Perspectives*, 112(1), pp.6–8. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1241116/ [Accessed December 14, 2014].

Bentayeb, M. et al., 2012. Adverse respiratory effects of outdoor air pollution in the elderly. *International Journal of Tuberculosis and Lung Disease*, 16(9), pp.1149–1161.

Bernstein, J. a. et al., 2004. Health effects of air pollution. *Journal of Allergy and Clinical Immunology*, 114, pp.1116–1123.

Van den Bossche, J. et al., 2014. Mapping spatial variation in urban air quality using mobile measurements: development and validation of a methodology based on an extensive dataset. *Atmospheric Environment*, submitted, pp.148–161. Available at: http://dx.doi.org/10.1016/j.atmosenv.2015.01.017.

Brunekreef, B. & Holgate, S.T., 2002. Air pollution and health. *Tidsskrift for den Norske laegeforening tidsskrift for praktisk medicin ny raekke*, 116(1), pp.2147–2148. Available at: http://dx.doi.org/10.1016/S0140-6736(02)11274-8.

Dons, E. et al., 2012. Personal exposure to Black Carbon in transport microenvironments. *Atmospheric Environment*, 55(July 2015), pp.392–398.

Dourish, P., 2004. What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8(1), pp.19–30.

EEA, 2014. *Air pollution fact sheet 2014*, Copenhagen.

EPA, 2015. Persistent Organic Pollutants: A Global Issue, A Global Response. *epa.org*.

EPA, 2012. Volatile Organic Compounds (VOCs). *epa.org*. Available at: http://www.epa.gov/iaq/voc.html [Accessed April 22, 2015].

EveryAware, 2014. *EveryAware White Paper*

Glass, R.L., 1995. A structure-based critique of contemporary computing research. *Journal of Systems and Software*, 28(1), pp.3–7.

Gordon, T. & Reibman, J., 2000. Cardiovascular toxicity of inhaled ambient particulate matter. *Toxicological sciences : an official journal of the Society of Toxicology*, 56, pp.2–4.

Greven, S., Dominici, F. & Zeger, S., 2011. An Approach to the Estimation of Chronic Air Pollution Effects Using Spatio-Temporal Information. *Journal of the American Statistical Association*, 106(February 2015), pp.396–406.

Hasenfratz, D. et al., 2014. Deriving High-Resolution Urban Air Pollution Maps Using Mobile Sensor Nodes. , pp.1–25.

Hasenfratz, D., Saukh, O. & Thiele, L., 2012. On-the-fly calibration of low-cost gas sensors. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7158 LNCS, pp.228–244.

Hertel, O. et al., 2008. 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, pp.58–70.

Høiskar, B.A.K., Sundvor, I. & Strand, A., 2014. *Tiltaksutredning for luftkvalitet i Oslo og Bærum 2015-2020 .* 01 ed.,

Jarjour, S. et al., 2013. Cyclist route choice, traffic-related air pollution, and lung function: a scripted exposure study. *Environmental Health: A Global Access Science Source*, 12(1), pp.1–12. Available at: 10.1186/1476-069X-12-14\nhttp://libproxy.wustl.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=8gh&AN=87481244&site=ehost-live&scope=site.

Jones, K.C. & de Voogt, P., 1999. Persistent organic pollutants (POPs): state of the science. *Environmental Pollution*, 100, pp.209–221. Available at: http://www.sciencedirect.com/science/article/B6VB5-3X3BXS6-F/2/0227124d55f9dfb4c8af97674f0566da.

Kampa, M. & Castanas, E., 2008. Human health effects of air pollution. *Environmental Pollution*, 151, pp.362–367.

Luftkvalitet.info, 2015. Luftforurensing. Available at: http://www.luftkvalitet.info/Theme.aspx?ThemeID=6fc2e3cd-424f-4c03-ad0c-2b9c15369cd9 [Accessed April 22, 2015].

MacNaughton, P. et al., 2014. Impact of bicycle route type on exposure to traffic-related air pollution. *Science of the Total Environment*, 490(2), pp.37–43. Available at: http://dx.doi.org/10.1016/j.scitotenv.2014.04.111.

Manouselis, N. et al., 2011. Recommender Systems in Technology Enhanced Learning. *Recommender Systems Handbook*, pp.387–415. Available at: http://www.springerlink.com/index/10.1007/978-0-387-85820-3.

De Nazelle, A. et al., 2012. A travel mode comparison of commuters' exposures to air pollutants in Barcelona. *Atmospheric Environment*, 59, pp.151–159. Available at: http://dx.doi.org/10.1016/j.atmosenv.2012.05.013.

Nguyen, T.T. et al., 2014. Exploring the Filter Bubble : The Effect of Using Recommender Systems on Content Diversity. *In Proceedings of the 23rd international conference on World wide web,*, pp.677–686. Available at: http://www-users.cs.umn.edu/~tien/publications/filter_bubble.pdf.

Pattinson, W., Longley, I. & Kingham, S., 2014. Using mobile monitoring to visualise diurnal variation of traffic pollutants across two near-highway neighbourhoods. *Atmospheric Environment*, 94, pp.782–792. Available at: http://dx.doi.org/10.1016/j.atmosenv.2014.06.007.

Peters, J. et al., 2014. Cyclist exposure to UFP and BC on urban routes in Antwerp, Belgium. *Atmospheric Environment*, 92, pp.31–43. Available at: http://dx.doi.org/10.1016/j.atmosenv.2014.03.039.

Pope, C.A., 2007. Mortality effects of longer term exposures to fine particulate air pollution: review of recent epidemiological evidence. *Inhalation toxicology*, 19 Suppl 1(October 2006), pp.33–38.

Pope, C.A. & Dockery, D.W., 2006. Health Effects of Fine Particulate Air Pollution: Lines that Connect. *Journal of the Air & Waste Management Association*, 56(February 2015), pp.709–742.

Resnick, P., Varian, H.R. & Editors, G., 1997. Recommender Systems. *Communications of the ACM*, 40(3), pp.56–58. Available at: http://doi.acm.org/10.1145/245108.245121.

Sarnat, J. a. et al., 2001. Gaseous pollutants in particulate matter epidemiology: Confounders or surrogates? *Environmental Health Perspectives*, 109(10), pp.1053–1061.

Schafer, J. Ben, Konstan, J. & Riedi, J., 1999. Recommender Systems in E-Commerce. *Proceedings of the 1st ACM conference on Electronic commerce EC 99*, 2001, pp.158–166. Available at: http://portal.acm.org/citation.cfm?doid=336992.337035.

Schilit, B.N. & Theimer, M.M., 1994. Disseminating Active Map Information to Mobile Hosts. *Network, IEEE*, 8(5), pp.22–32.

Seaton, a et al., 1995. Particulate air pollution and acute health effects. *Lancet*, 345(8943), pp.176–178.

Snyder, E.G. et al., 2013. The changing paradigm of air pollution monitoring. *Environmental Science & Technology*, 47(20), pp.11369–11377. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23980922\nhttp://dx.doi.org/10.1021/es4022602.

Soghoian, S., 2009. Heavy Metal Toxicity. *medisuite.ir*. Available at: http://medisuite.ir/medscape/a814960-business.html#a01 [Accessed January 31, 2015].

Srivastava, A. & Rao, B.P.S., 2011. Urban Air Pollution Modeling. In P. D. Popovic, ed. *Air Quality-Models and Applications*. p. 364.

Tchounwou, P.B. et al., 2012. Heavy Metals Toxicity and the Environment. *Molecular, Clinical and Environmental Toxicology*, 101, pp.133–164. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4144270/.

Thai, A., McKendry, I. & Brauer, M., 2008. Particulate matter exposure along designated bicycle routes in Vancouver, British Columbia. *Science of the Total Environment*, 405, pp.26–35.

Tintarev, N. & Masthoff, J., 2011. *Recommender Systems Handbook*, Available at: http://www.springerlink.com/index/10.1007/978-0-387-85820-3.

Verbert, K. et al., 2012. Context-aware recommender systems for learning: A survey and future challenges. *IEEE Transactions on Learning Technologies*, 5(4), pp.318–335.

Werthner, H., Hansen, H.R. & Ricci, F., 2007. The Big Promise of Recommender Systems. *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*, pp.19–27.

World Health Organization, 2013. Health Effects of Ambient Particulate Matter. *Journal of the Korean Medical Association*, 50, p.175. Available at: http://synapse.koreamed.org/DOIx.php?id=10.5124/jkma.2007.50.2.175.

Zhang, Y. & Wang, L., 2010. Some challenges for context-aware recommender systems. *ICCSE 2010 - 5th International Conference on Computer Science and Education, Final Program and Book of Abstracts*, pp.362–365.

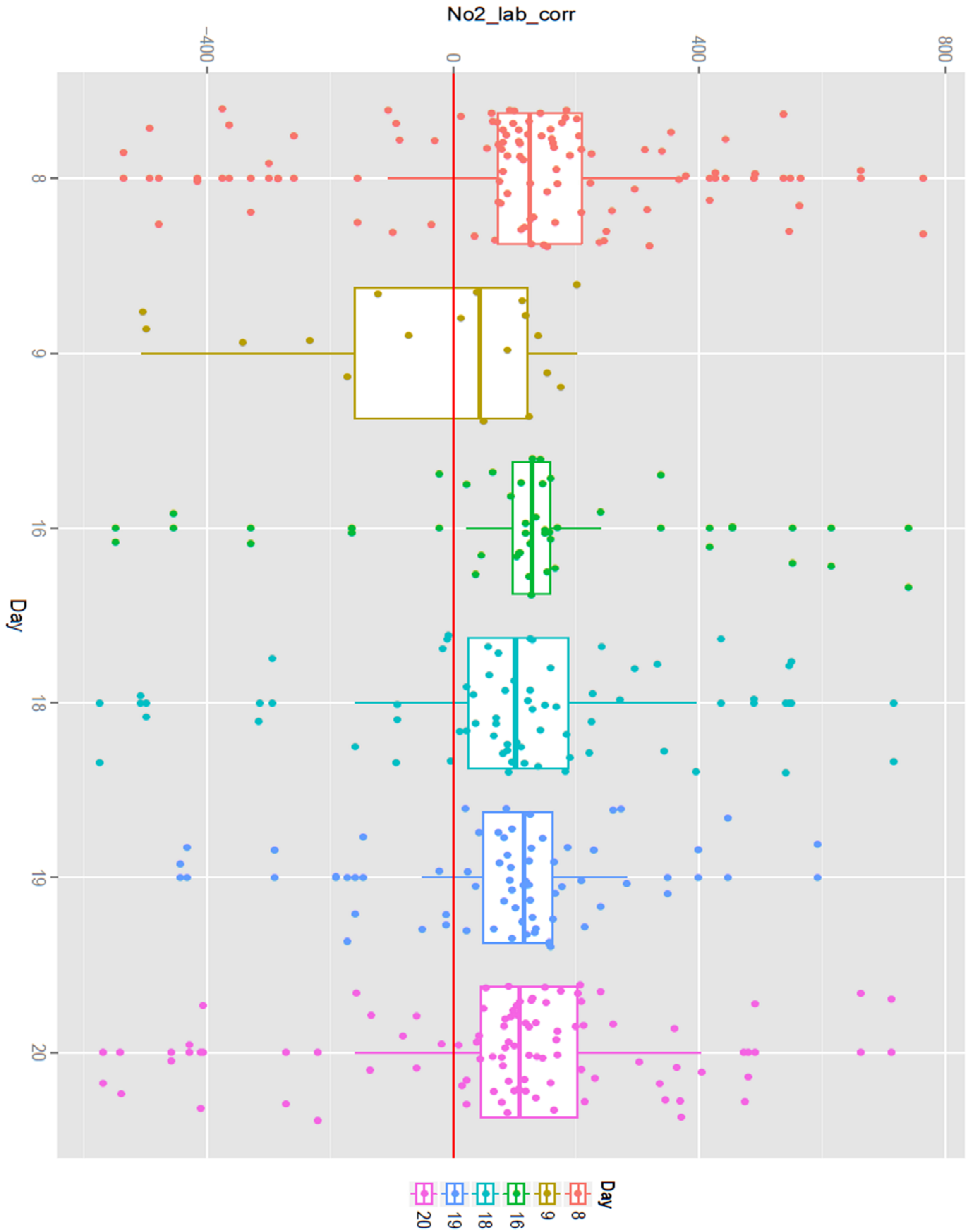
Zuurbier, M. et al., 2010. Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route. *Environmental Health Perspectives*, 118(6), pp.783–789.

Zwack, L.M. et al., 2011. Modeling spatial patterns of traffic-related air pollutants in complex urban terrain. *Environmental Health Perspectives*, 119(6), pp.852–859.

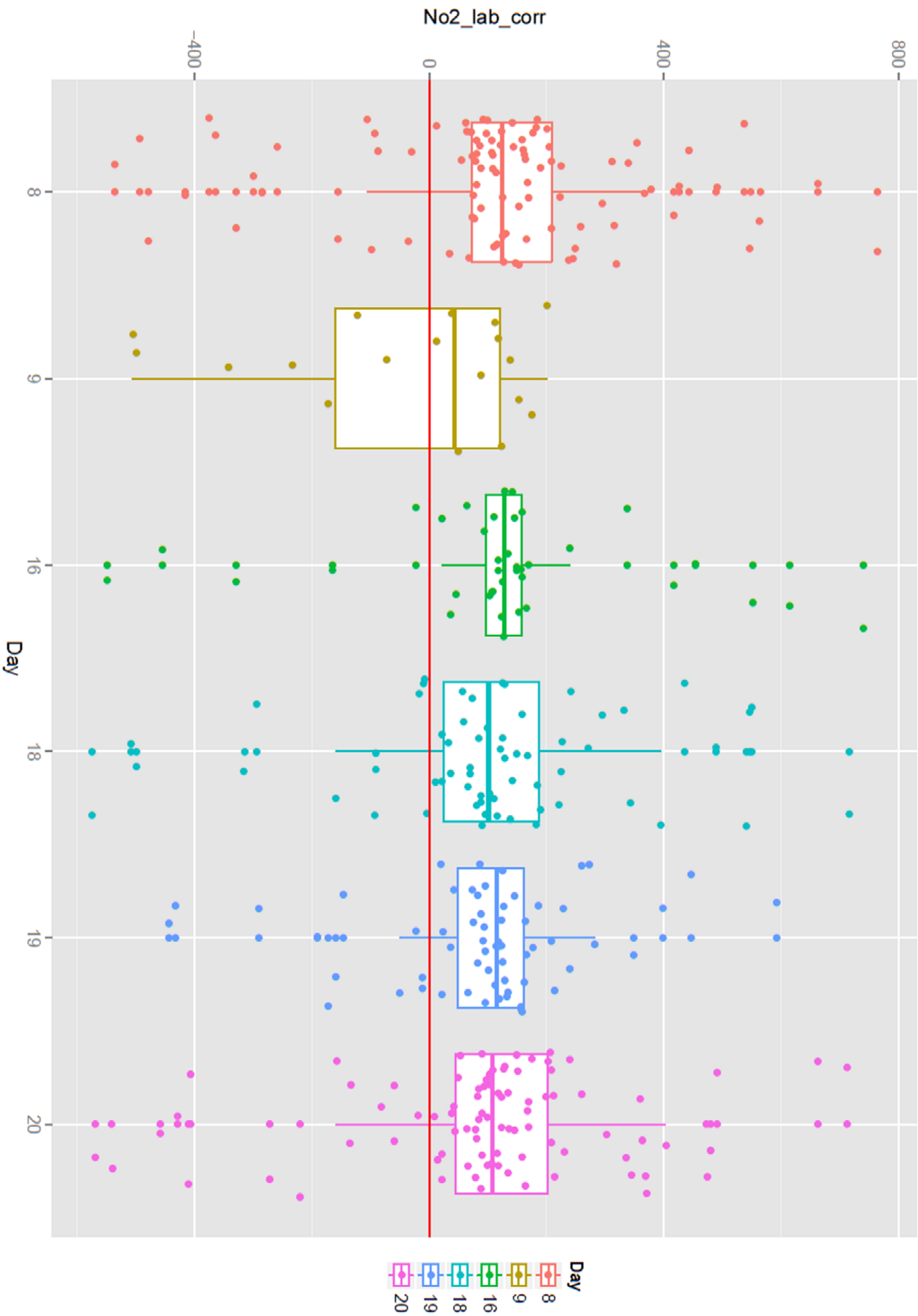
Appendix A

Boxplots of lab corrected values from the Kirkeveien route.

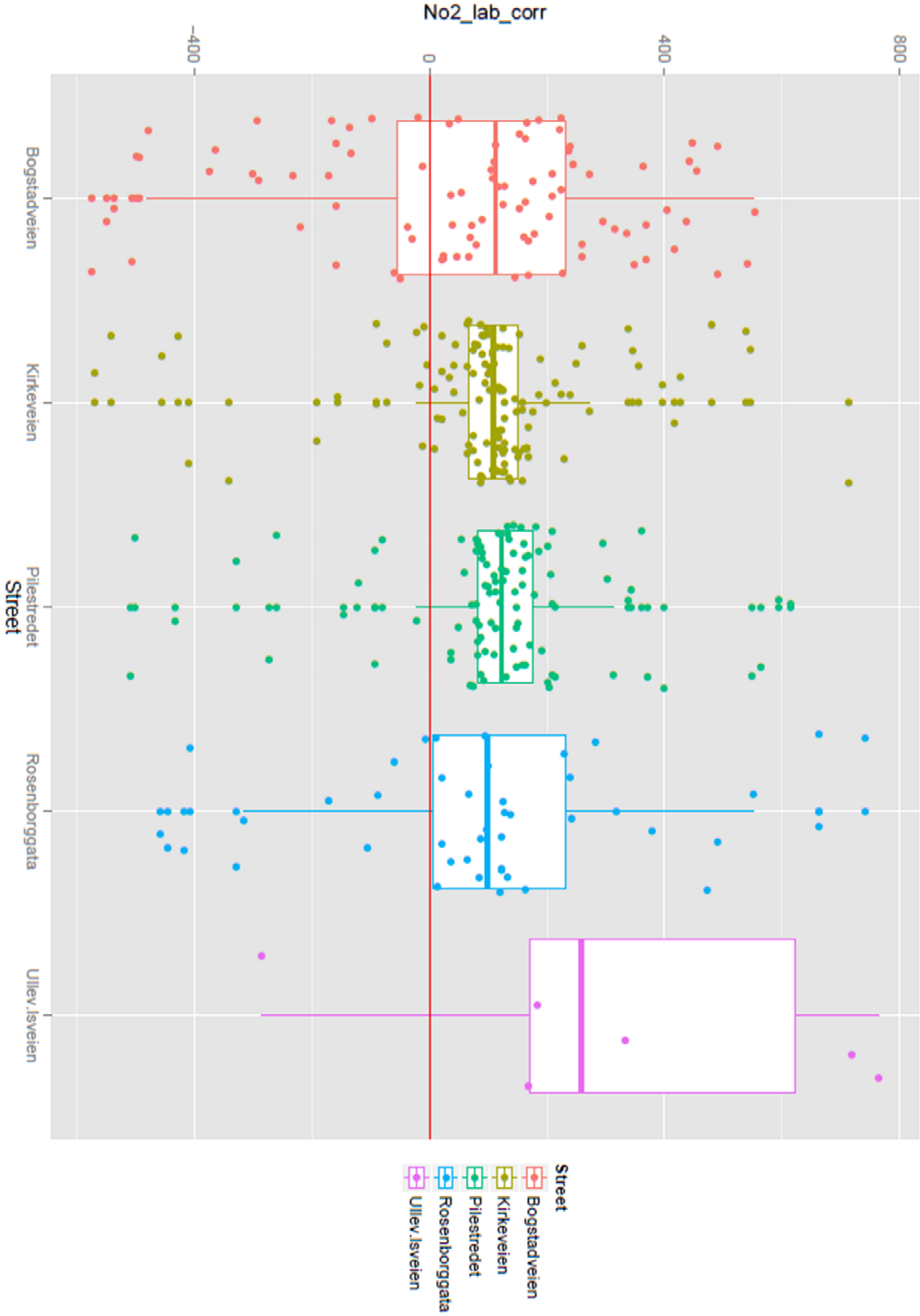
1) Day-to-day variations in NO_2 levels:



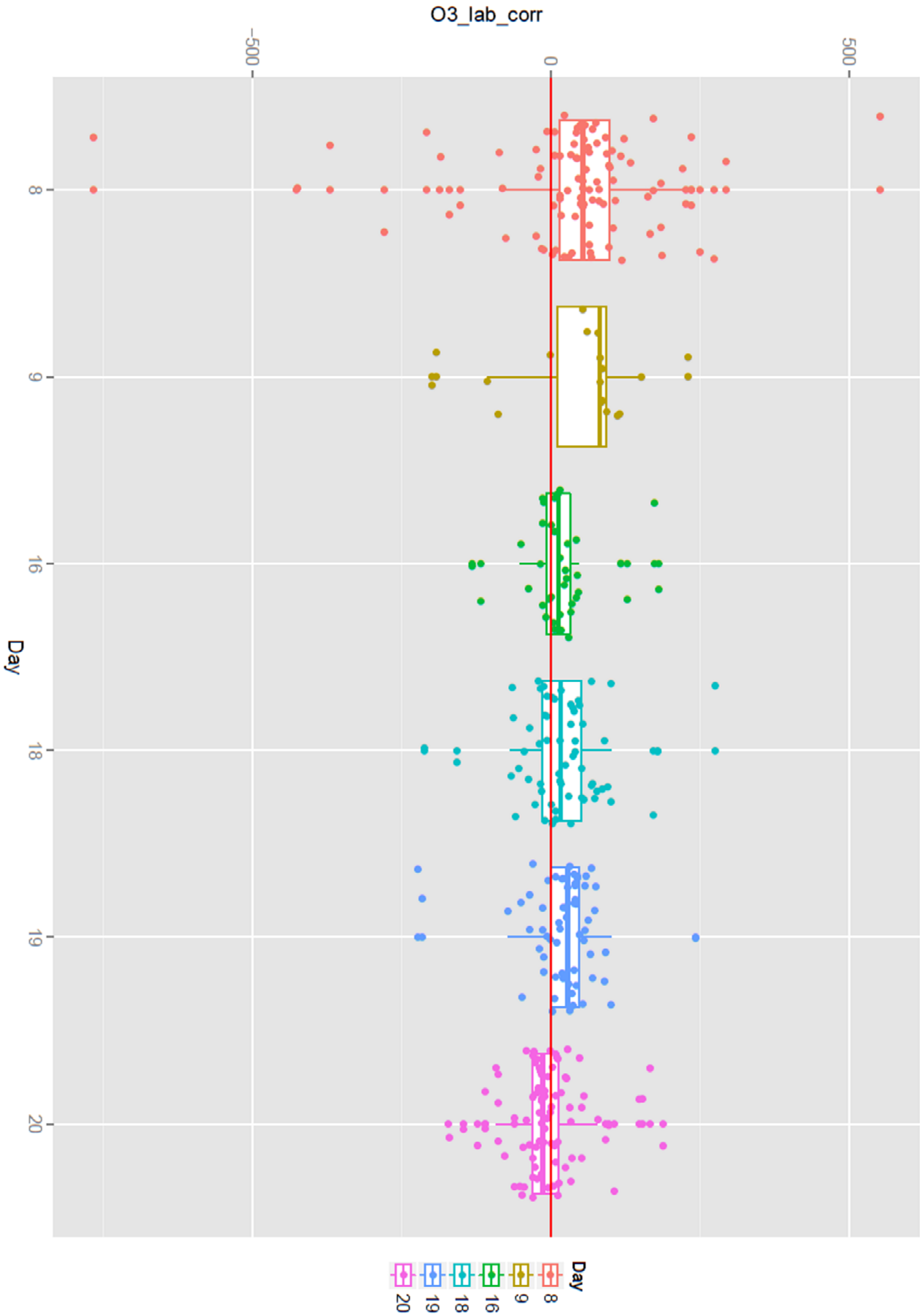
2) Hourly variations in NO_2 levels:



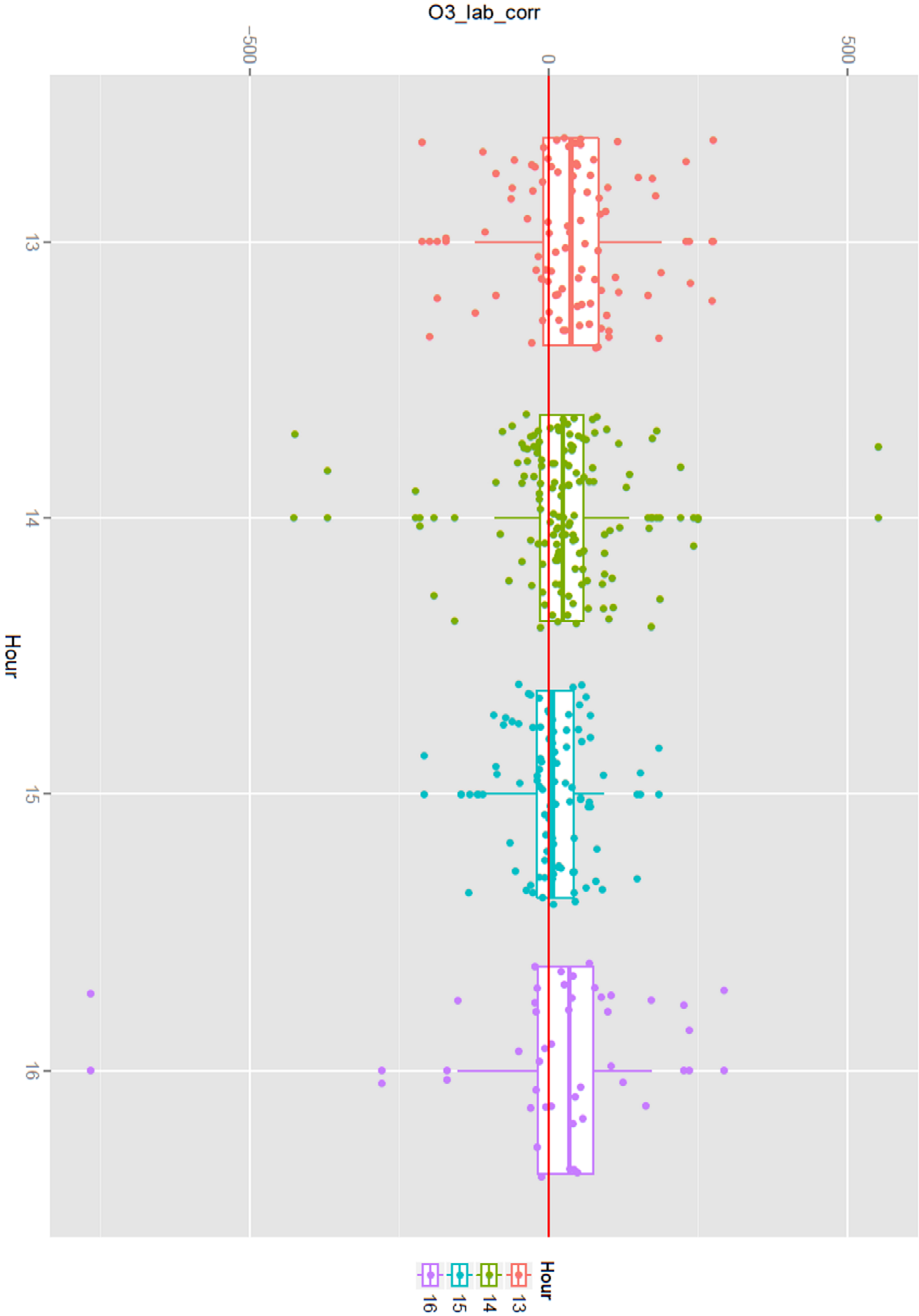
3) Street-to-street variations in NO_2 levels



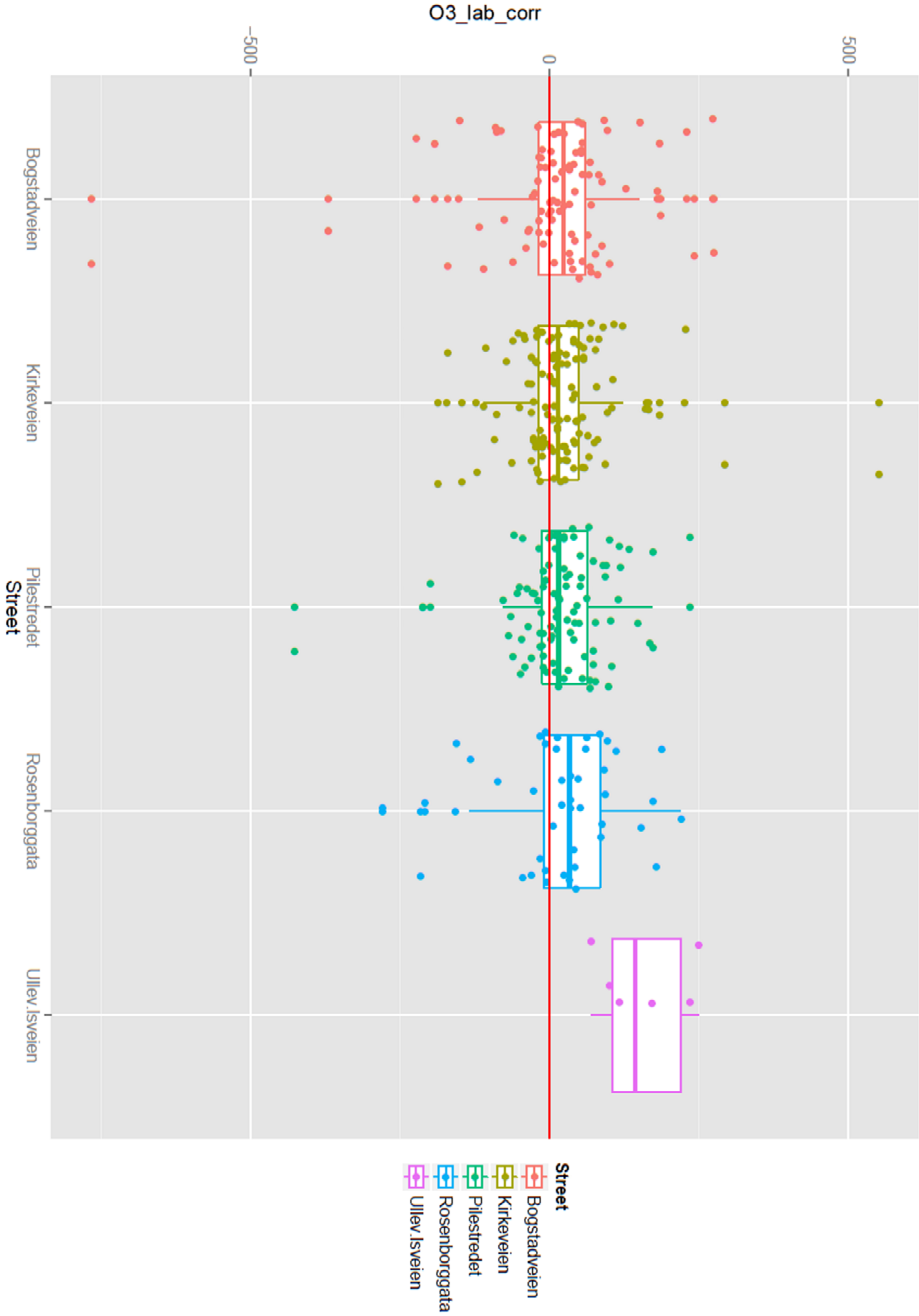
4) Day-to-day variations in O_3 levels, 4 outliers has been removed for readability



5) Hourly variations in O_3 levels, 4 outliers has been removed for readability



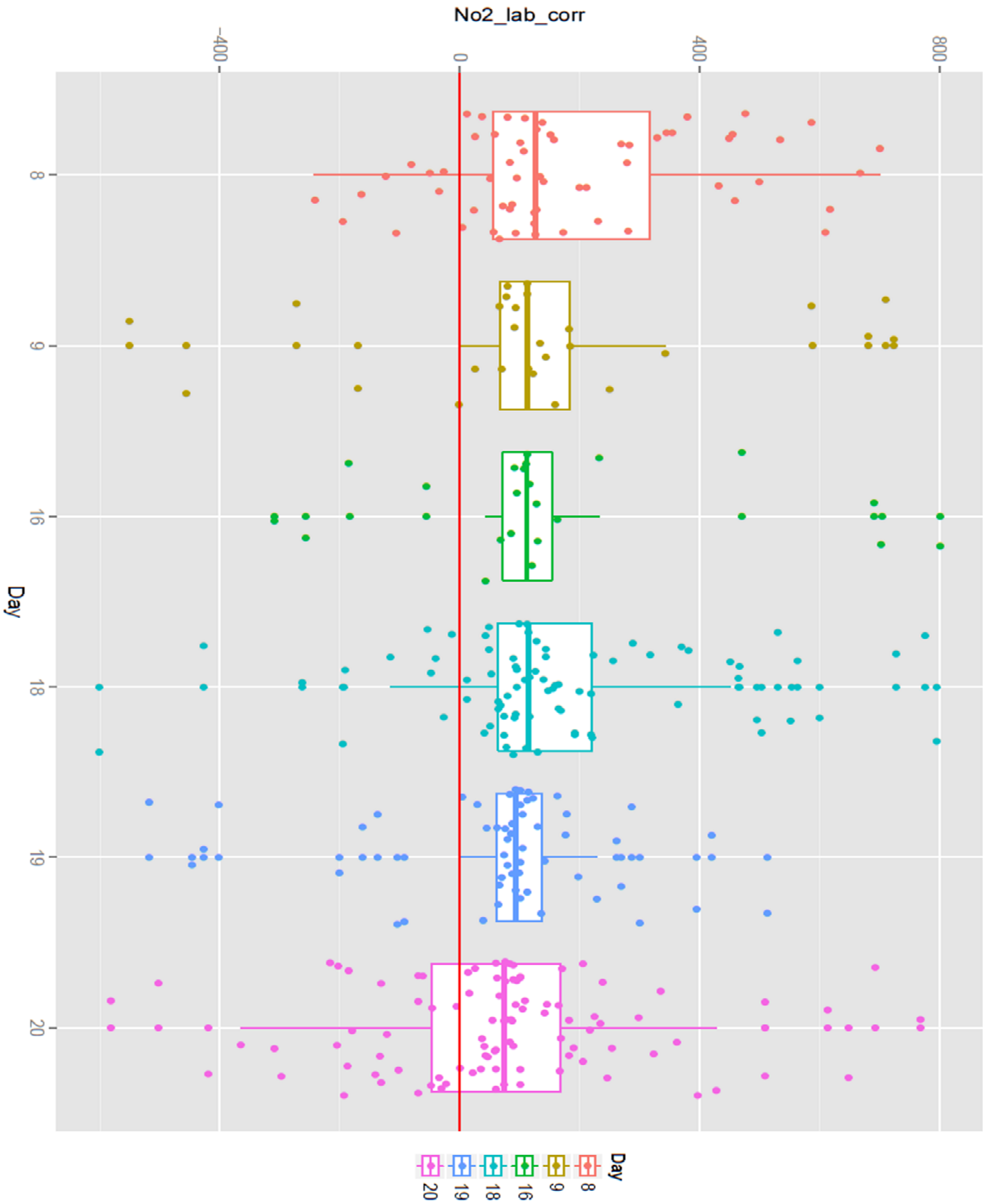
6) Street-to-street variations in O_3 levels, 4 outliers has been removed for readability



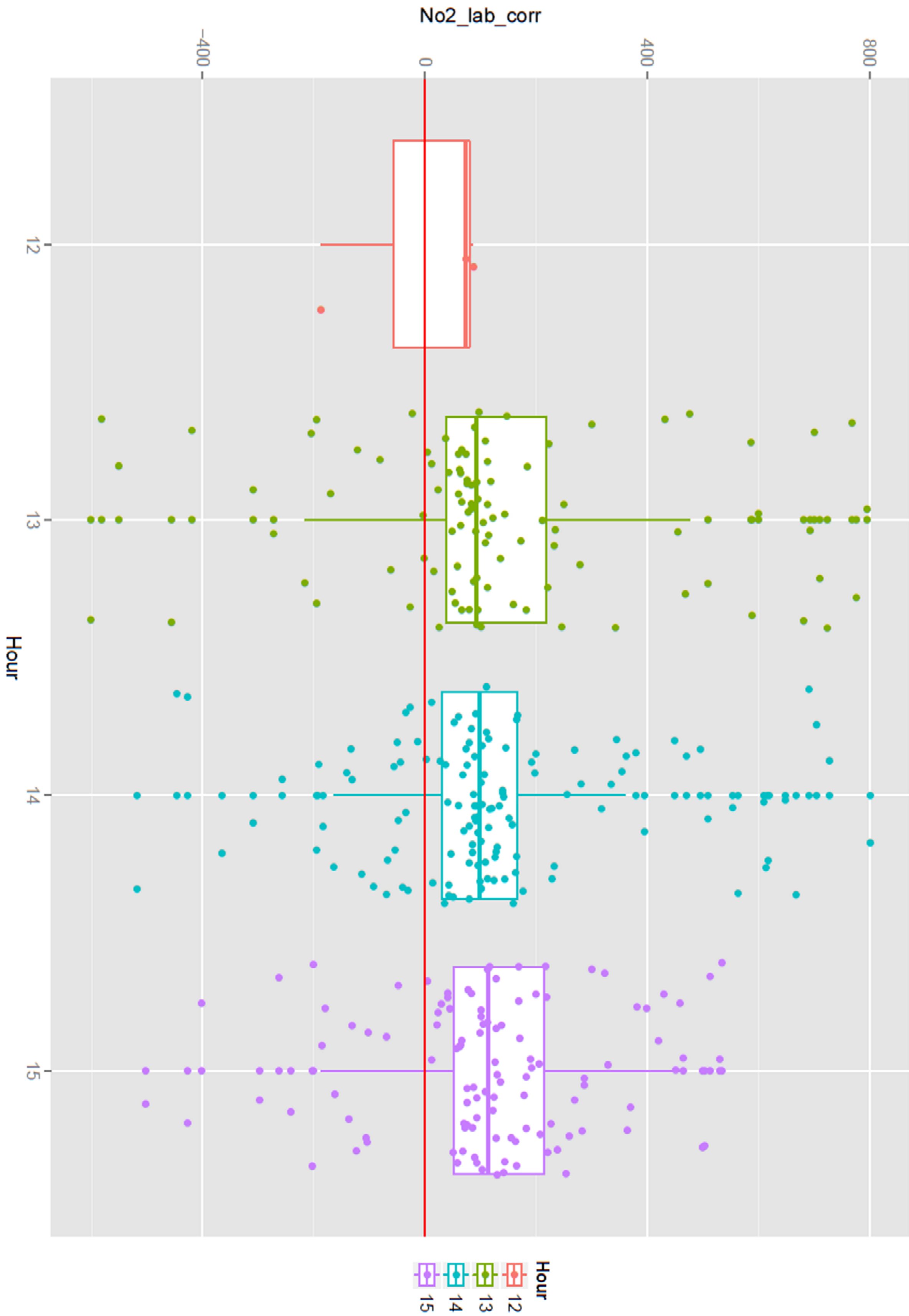
Appendix B

Boxplots of lab corrected values from the Fagerborg route.

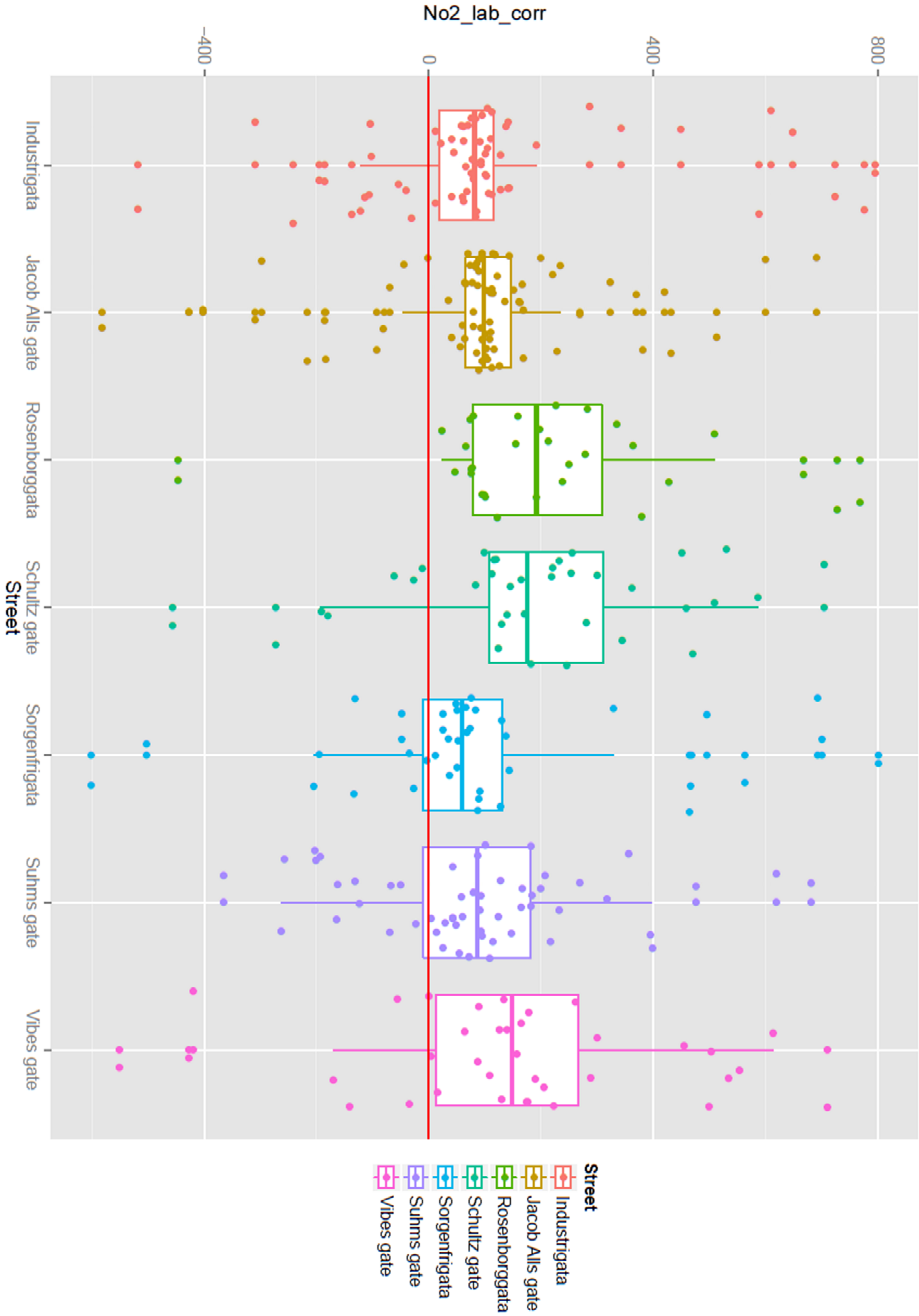
1) Day-to-day variations in NO_2 levels:



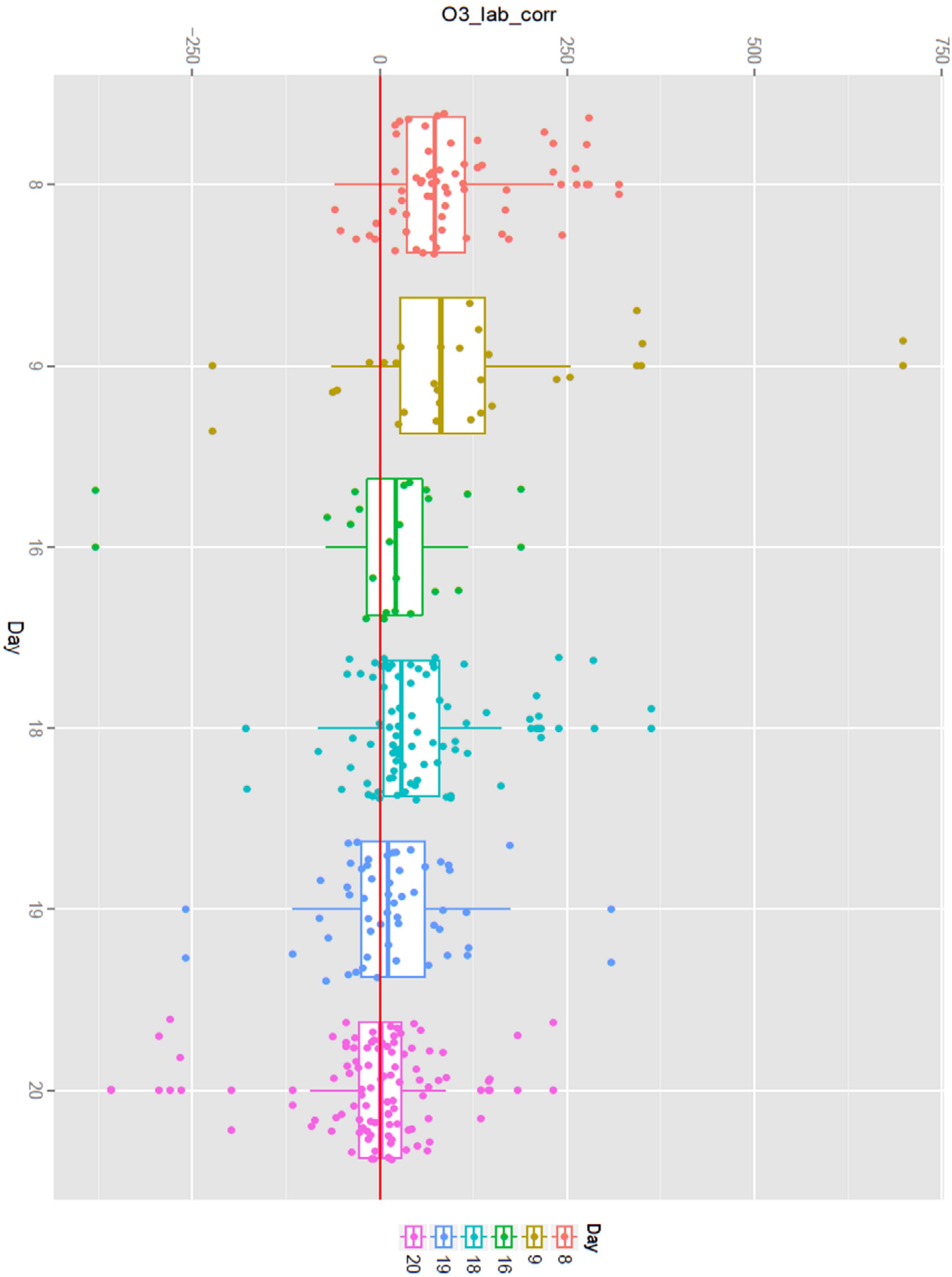
2) Hourly variations in NO_2 levels:



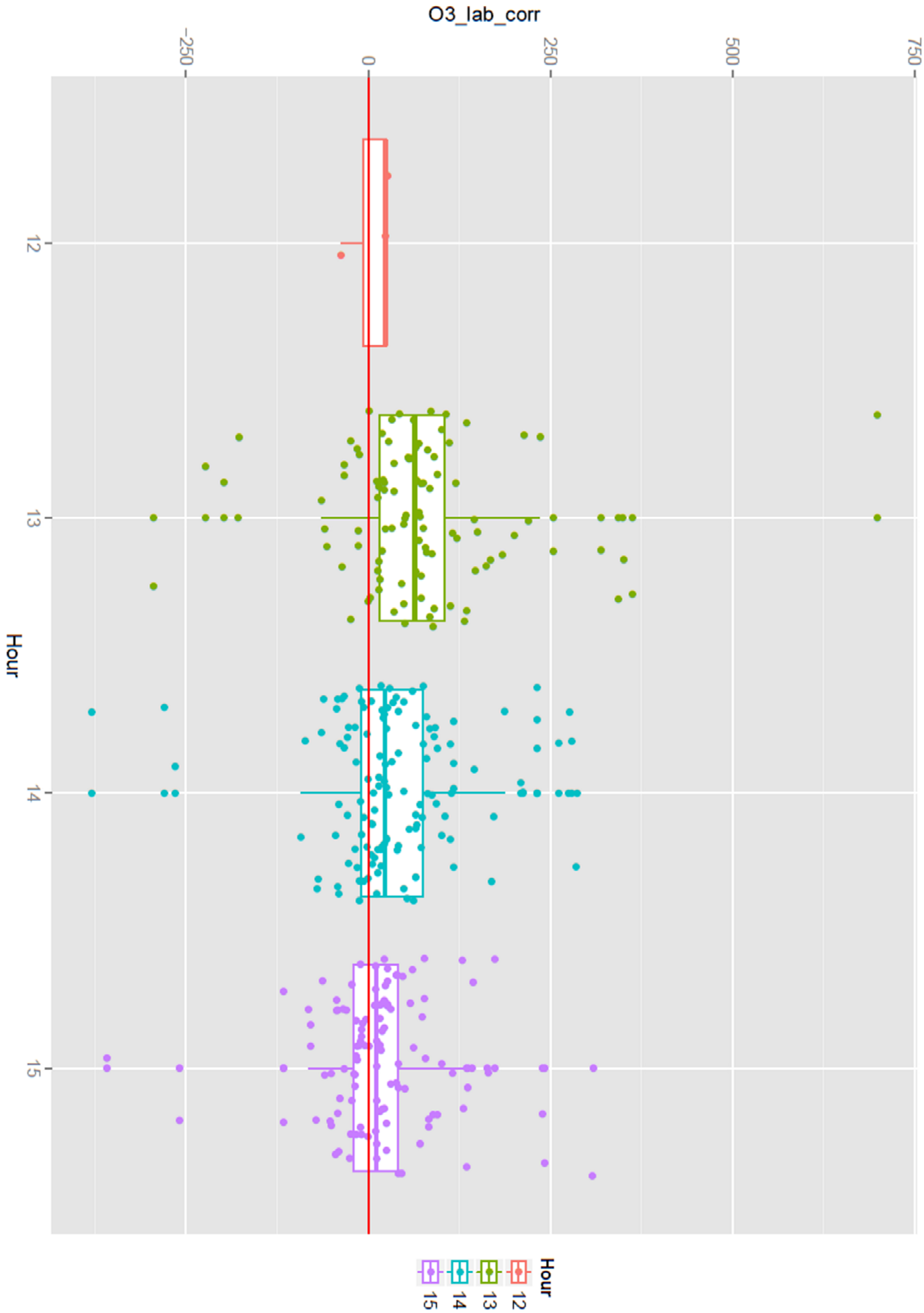
3) Street-to-street variations in NO_2 levels



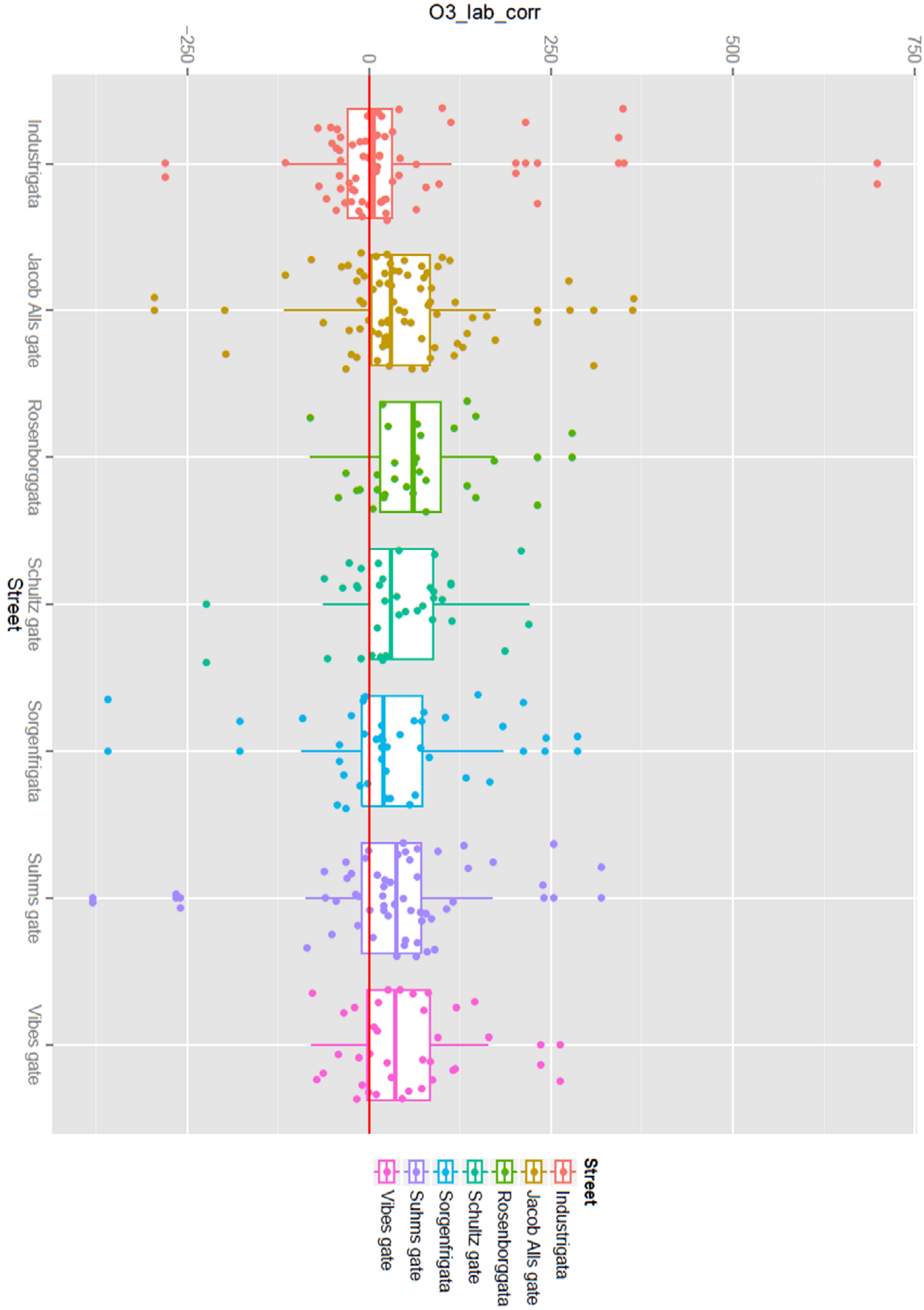
4) Day-to-day variations in O_3 levels



5) Hourly variations in O_3 levels



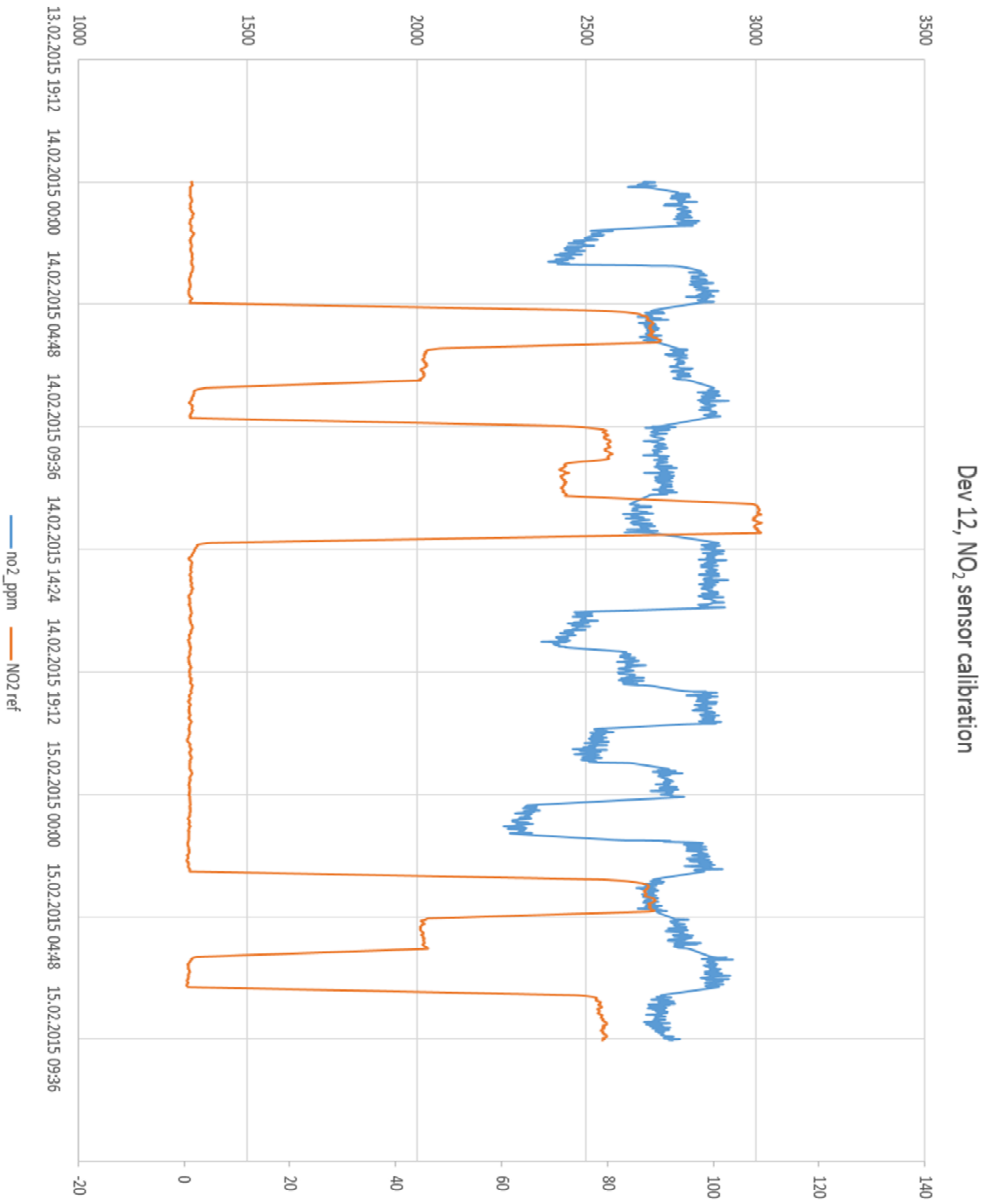
6) Street-to-street variations in O_3 level



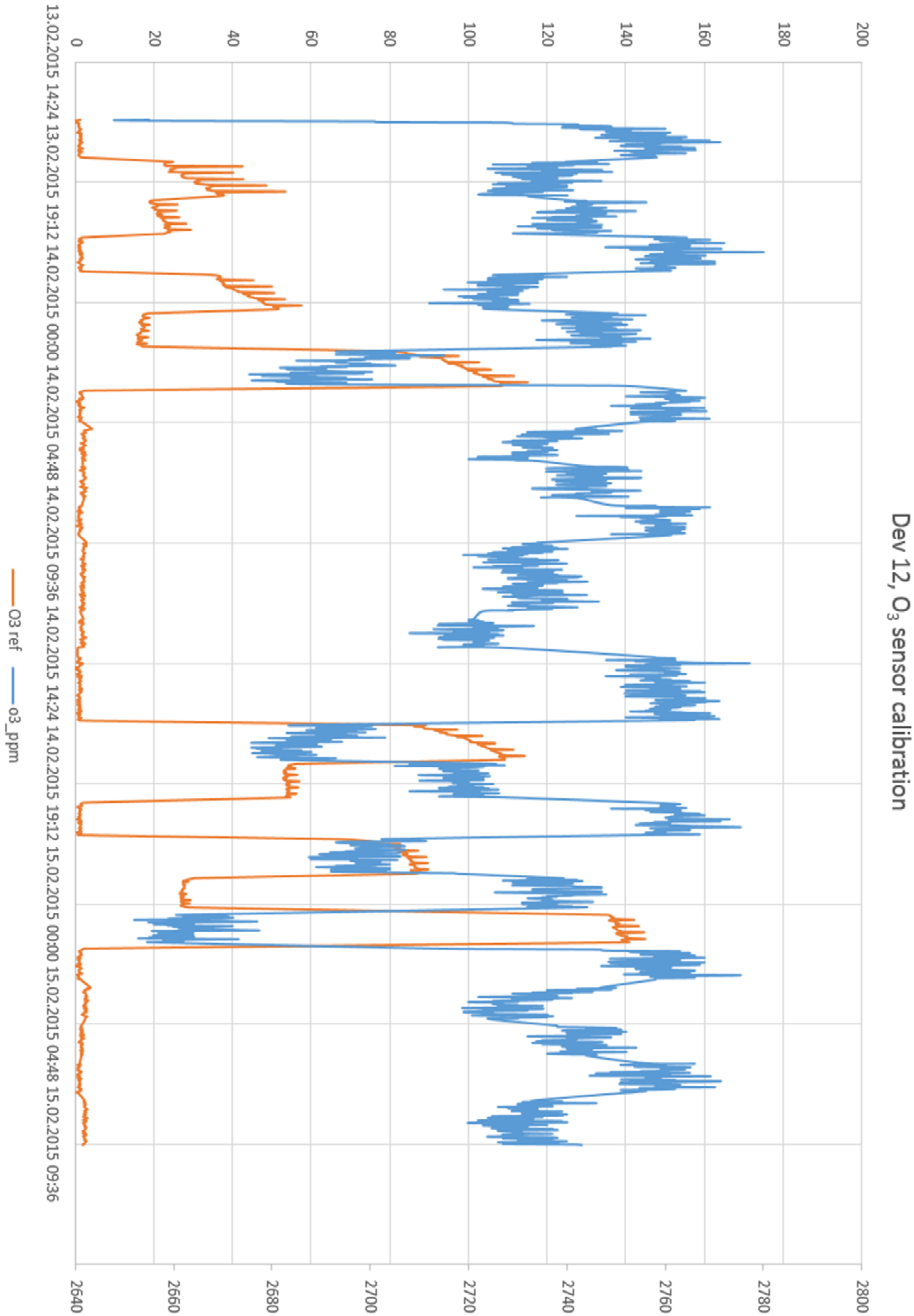
Appendix C

Calibration and correlation results from NILU

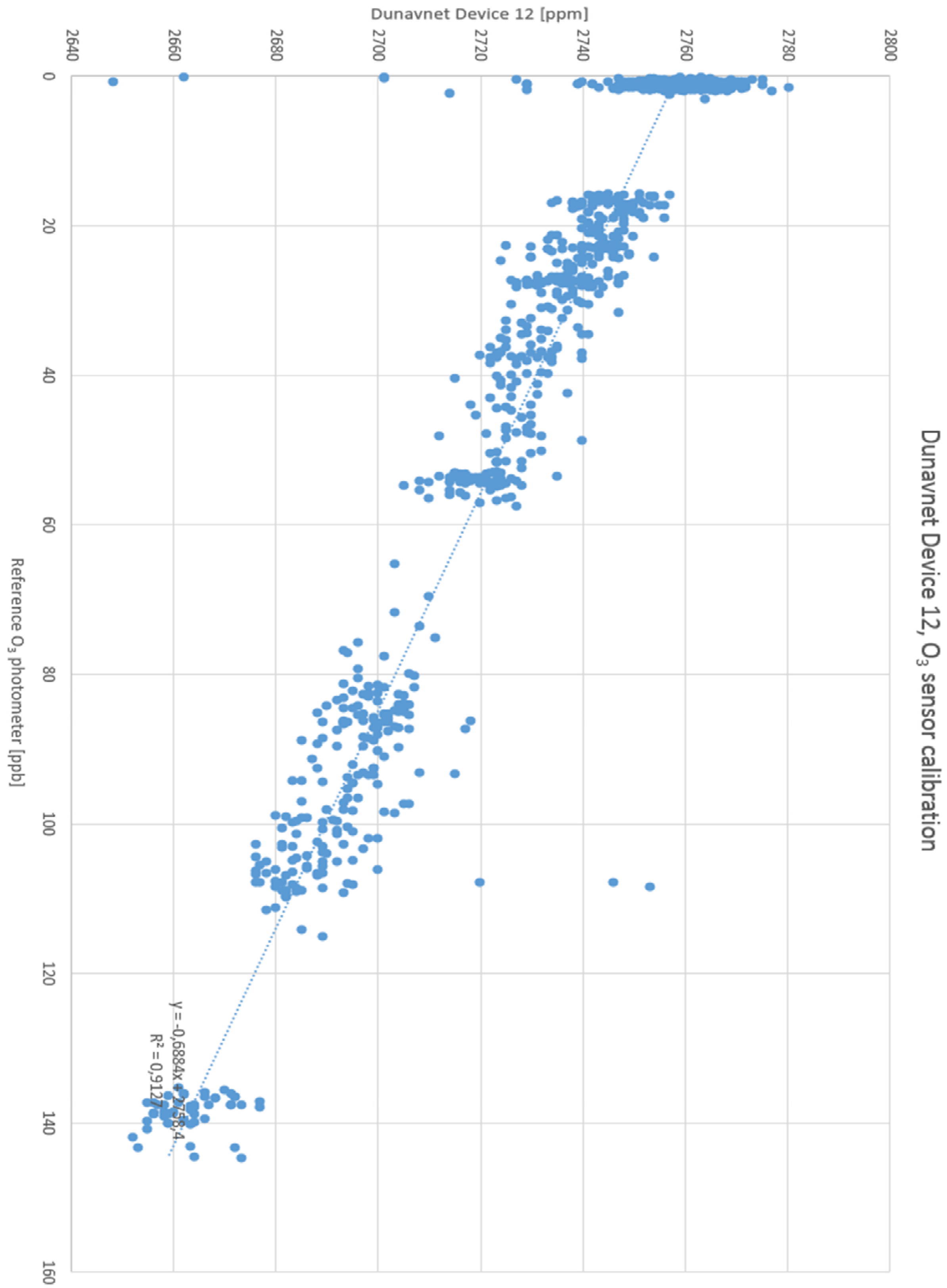
NO_2 sensor calibration:



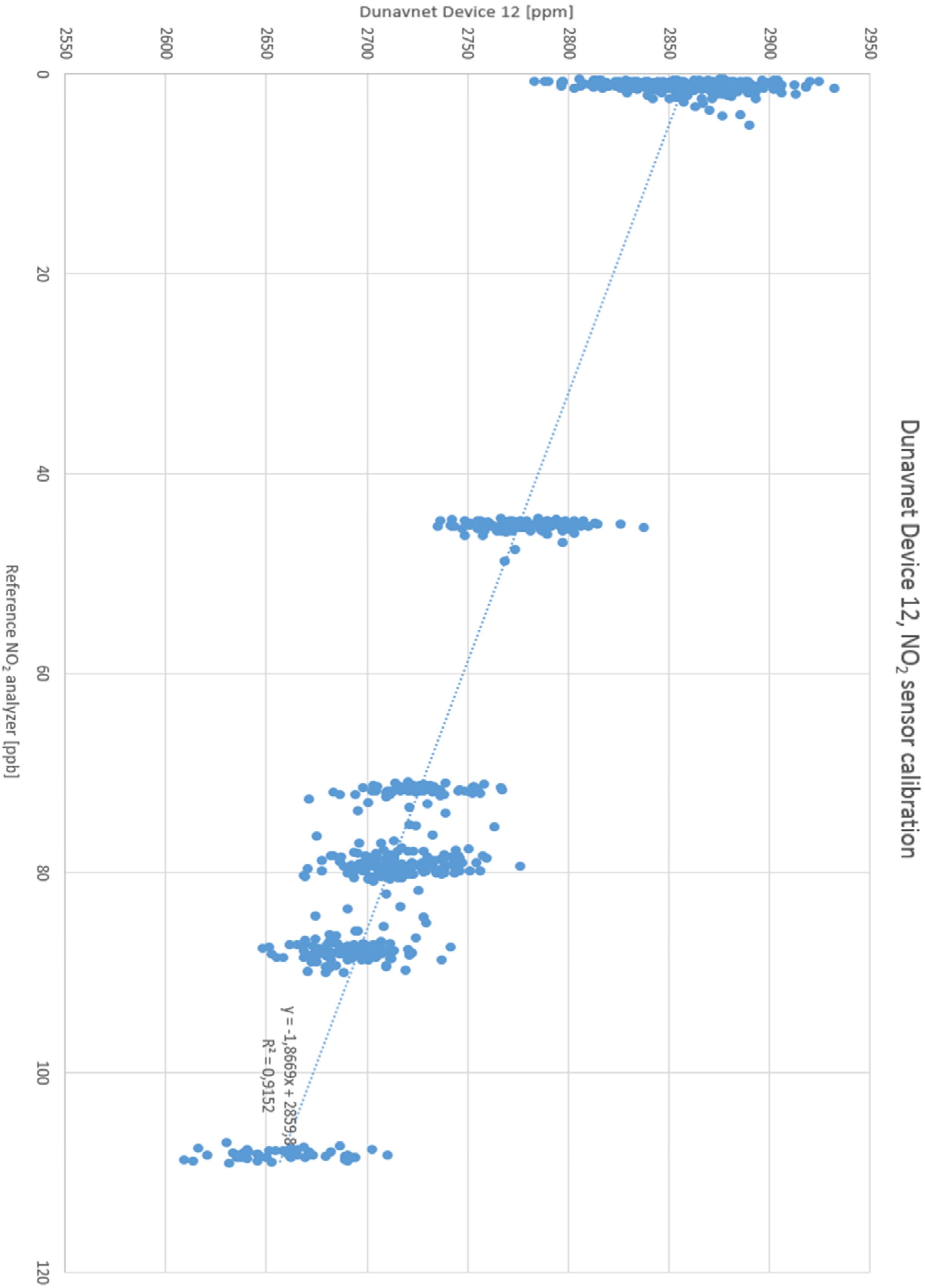
O_3 sensor calibration:



O$_3$ Correlation chart:



NO_2 Correlation chart:



Appendix D

The datasets used in this thesis are too large to append, so they are made available to view and download from figshare.com.

Raw dataset from the monitoring platform for all the days:

figshare.com/s/d3713b28393711e59ee506ec4b8d1f61

Cleaned dataset from the Kirkeveien route:

figshare.com/s/c5f2bd6e393711e5b3c706ec4b8d1f61

Cleaned dataset from the Fagerborg route:

figshare.com/s/10c53aac393711e5ab5f06ec4b8d1f61