

**Sensing Spaces: Personal
Exposure to Air Pollution with
Different Modes of Transport and
Urban Environments Using
Wearable Sensors**

Mihai Visuian

4th Year Project Report
Computer Science
School of Informatics
University of Edinburgh

2018

Abstract

This is an example of `infthesis` style. The file `skeleton.tex` generates this document and can be used to get a “skeleton” for your thesis. The abstract should summarise your report and fit in the space on the first page. You may, of course, use any other software to write your report, as long as you follow the same style. That means: producing a title page as given here, and including a table of contents and bibliography.

Acknowledgements

I would like to thank my supervisor, professor D K Arvind for his main support in helping me work thoroughly on a project of such high complexity and importance in the real world. I would also like to thank him for his valuable advice and guidance.

Furthermore, I would like to thank Andrew Bates for his support in providing equipment and helping with maintenance of sensors and servers.

Table of Contents

1	Introduction	7
1.1	Motivation	7
1.2	Objectives	7
1.3	Literature Review	8
1.4	Method	9
1.5	Structure of the Report	10
2	State of the Art	11
3	Methodology	13
4	Implementation	15
4.1	Data Visualisation Tool	15
5	Results	17
6	Conclusions	19
6.1	Concluding Remarks	19
6.2	Future Work	19
	Bibliography	21

Chapter 1

Introduction

1.1 Motivation

Air pollution in large urban environments is a controversial topic nowadays and pollution sources, as well as distribution and effect on human health have been widely studied. Particulate matter (PM) represents the sum of all solid particles and liquid droplets of small sizes that are suspended in the air. It has been discovered that exposure to such particles of size less than 10 microns can be hazardous for the human body and it has been associated with increased mortality rates in areas labelled as highly polluted [5].

Air quality is mostly measured through stationary Air Quality Monitoring Stations (AQMS). They measure air quality at a fixed location on a certain radius with high precision. Data collected from such networks of stations distributed strategically within urban environments is used to determine exposure to pollutants and is used for checking if a specific area meets requirements set by legislation. One drawback of this approach is that monitoring a large fixed region through a stationary sensor would not always suffice to obtain accurate information about air quality and pollution sources, as parameters vary from a corner of the region to another and other external factors such as wind speed and terrain type affect the results.

One approach to solve this issue and obtain more accurate data from the environment would be to deploy low-cost mobile sensors on people, such as drivers, pedestrians and cyclists in order to obtain a more precise spatial representation of the air quality. However, these sensors have to be calibrated in correlation with the Air Quality Monitoring Stations to ensure high precision of the information gathered.

1.2 Objectives

This project mainly focuses on Mobile Exposure Monitors (MEM) developed by the Centre for Speckled Computing. These wearable devices are equipped with sensors measuring humidity, temperature and particulate matter. Data is transmitted to an An-

droid device via Bluetooth Low Energy. The project presents several study methods for detecting urban environments and personal exposure with different means of transport, based on air quality data collected with the wearable sensors. The objectives of the project are:

- Collect data at specific times on a daily basis and on a fixed route.
- Perform analysis on data gathered for various modes of transport and urban environments and research machine learning techniques to detect personal exposure on a specific mode of transport or urban environment based on temperature, humidity and particulate matter attributes.
- Build a data visualization tool for a more accurate spatial representation of data gathered and results obtained after applying statistical and machine learning methods.
- Validate measurements of sensors against a more accurate Air Quality Monitoring Station in Edinburgh.

1.3 Literature Review

There are past studies concerning exposure of cyclists to air pollution and traffic noise in central neighbourhoods of Montreal [3]. Despite it is known that cycling in urban areas would have beneficial effects on people's health, it is associated with potential health risks due to long time exposure to polluted air and traffic noise. The focus of this study was to analyse exposure of cyclists to air pollution and traffic noise and detect the impact on exposure of local factors such as weather conditions and the type of roads. A number of 85 bicycle trips were analysed, totalling 422 km of travel. The results revealed a weak negative correlation between noise and air pollution (NO₂) measures of exposure.

A study [2] done in Auckland, New Zealand, examined personal exposure to air pollution, namely carbon monoxide (CO) for several modes of transport, using Langan T15n [1] portable carbon monoxide monitors. Participants were constrained to travel at specific peak traffic times between designated start and end points. Results suggest that lowest exposure to CO particles are experienced by train commuters, while motorcyclists were exposed to significantly higher average concentrations. Furthermore, travel by bus on a dedicated route proved to be more effective than travel by car on a congested motorway. Also, average exposure of cyclists and pedestrians proved to be similar to bus and car commuters. However, when increased physical activity that implies higher volumes of air breathed, along with increased commuting time were taken into account, air pollution dosage became much higher than for the motorised means of transport.

Most research regarding particulate matter has been performed in terms of values of PM_{2.5} and PM₁₀, namely measurements of mass particulates with lower dimensions than 2.5 and 10 microns. A study [4] is concerned about the relationship between mortality rate and particulate matter measured in Seoul, South Korea by an Optical Partic-

ulate Counter (OPC) and a national monitoring station. Particulate matter (PM_{2.5} and PM₁₀) and particle counts from 0.3 to 25 microns were measured. The results concerning particle counts were associated with a 5.73% and 5.82% in mortality caused by respiratory diseases.

1.4 Method

Data collection has been performed in several phases, using a monitoring device worn around the chest, measuring particulate matter (PM₁, PM_{2.5} and PM₁₀ values along with particle counts of different sizes), as well as temperature and humidity. The sensor communicates with an Android device via Bluetooth to which the data gathered is sent. First, data was gathered on Nicholson Street while using a taxi to Calton Hill during midday. Then, data was gathered while walking on the same route, at the same time of the day, in order to perform the first classification analysis between different means of transport (walking as opposed to cars). For further investigation of personal exposure with different modes of transport, another set of data collected was obtained from a train trip from Edinburgh to London King's Cross, including a bus trip to Edinburgh Waverley, as well as walking trips in Edinburgh, London and London Underground.

Data analysis has been performed using Jupyter Notebooks. First, the raw data was labelled accordingly in terms of the mean of transport used at each point in time. Secondly, the raw data was checked for any outliers existing in all the values of interest. After removing all irrelevant outliers, several machine learning classifiers have been used on PM data, along with temperature and humidity, in order to perform a classification of means of transport used during personal exposure experiments.

As far as personal exposure with different urban environments is concerned, a data collection schedule has been organised on a route in Edinburgh, including Nicholson Street, Melville Drive, Middle Meadow Walk, Tollcross and Lauriston Place, so that all types of urban environments would be detected. More exactly, the same route was covered over a week at two different times of the day, first at midday, during lunchtime between 13:00 and 14:00 and then in the evening, between 17:00 and 18:00, when a traffic peak hour is most probable to occur.

As in the previous set of experiments, Jupyter Notebooks have been used for analysis of the dataset regarding different urban environments. In this case, unsupervised machine learning methods have been applied for classification of data, as ranges of PM values corresponding to each urban environment type vary from a region to another. Thus, K-means clustering was performed first on the coordinates of the dataset features, such that all samples would be grouped into compact clusters and the mean values of the PM related values would be computed for each cluster. Then, a second K-means clustering was applied to the mean values of the clusters obtained in the previous step in order to detect the different types of urban environments.

1.5 Structure of the Report

Chapter 2 reveals detailed information regarding equipments used in the data collection phase, as well as details about PM measurements obtained. Moreover, details about the modes of transport used and types of urban environments taken into consideration are presented.

Chapter 3 details the data collection phase. Then, it also introduces details about statistics and machine learning techniques used in the data analysis process. Moreover, it introduces a data visualization tool implemented from scratch to ease analysis of data collected. Details about this tool will be further discussed in the following chapter.

Chapter 4 details on the data visualization tool, focusing both on the purpose and the functionality of the tool. The application is formed of two main parts, namely the front-end dealing with the spatial representation of the data on a map and the back-end, which consists of an API making requests to the data collected which is stored in a database.

Chapter 5 presents the results obtained from the data analysis phase. First, a research regarding classification of pollution exposure with different modes of transport is done using various machine learning classifiers. Then, unsupervised data is used to detect several types of urban environments, which will be presented in detail in chapter 2.

Finally, chapter 6 presents the conclusions drawn from the results obtained from data analysis and described in detail in chapter 5. Moreover, potential improvements and further study areas regarding the main topic of the project worth exploring are described.

Chapter 2

State of the Art

Chapter 3

Methodology

Chapter 4

Implementation

4.1 Data Visualisation Tool

Chapter 5

Results

Chapter 6

Conclusions

6.1 Concluding Remarks

6.2 Future Work

Bibliography

- [1] Langan T15n specification. <http://www.langan.biz/Site/T15n> Accessed: 2018-01-26.
- [2] Dirks , Sharma , Jennifer Salmond, Costello , and SB . Personal exposure to air pollution for various modes of transport in auckland, new zealand. 6 (Suppl 1:M5):84–92, 01 2012.
- [3] Philippe Apparicio, Mathieu Carrier, Jrmy Gelb, Anne-Marie Sguin, and Simon Kingham. Cyclists’ exposure to air pollution and road traffic noise in central city neighbourhoods of montreal. *Journal of Transport Geography*, 57:63 – 69, 2016.
- [4] Yong-Sung Cho, Jong-Tae Lee, Chang Jung, Young-Sin Chun, and Yoon-Shin Kim. Relationship between particulate matter measured by optical particle counter and mortality in seoul, korea, during 2001. 71:37–43, 10 2008.
- [5] Douglas Dockery, C Pope, Xuebing Xu, Jack Spengler, James H. Ware, Martha E. Fay, Benjamin G. Ferris, and Frank Speizer. An association between air pollution and mortality in six u.s. cities. 329:1753–9, 01 1994.