DEPARTMENT OF GEOGRAPHY

BA/BSc Geography

Please complete the following declaration and hand this form in with your dissertation

I, Jakub Wyszomierski

hereby declare :

(a) that this dissertation is my own original work and that all source material used is acknowledged therein;

(b) that it has been prepared specially for the BA/BSc in Geography of University College London;

(c) that it does not contain any material previously submitted to the Examiners of this or any other University, or any material previously submitted for any other examination.

Number of words: 11 748

Signed :

Date : 15/03/2019

A Quantitative Analysis of Running Routes Taken by Strava Users in London, and Their Exposure to Air Pollution.

By

Jakub Wyszomierski

2019

BA Geography with Quantitative Methods

Department of Geography

Faculty of Social and Historical Sciences

University College London (UCL)

Supervisors:

Dr Mirco Musolesi

Dr James Cheshire

Abstract

With a raising concern about health and environmental implications of air pollution, there is an increasing public interest in assessing and limiting the exposure to high levels of air pollution. Nonetheless, only a few studies have examined real-time mobility in the context of long-term exposure, which is seen as more important and short-term exposure.

Utilising data on runs completed by Strava users and data on modelled annual average air pollution from London Atmospheric Emissions Inventory, this panel study aims to conduct a quantitative analysis of the running routes and the associated exposure to air pollution.

In order to calculate the time-integrated exposure rate, the study examines ambient annual average concentration of $PM_{2.5}$ and predicts the intake of the pollutant, based on the duration of the exercise and changes to the ventilation rate caused by the variation in inclination.

The results suggest that the duration of the activity has a greater influence on the time-integrated exposure rate than variations in the ventilation rate. It is suggested that roads characterized by lower air pollution intake, but significantly longer duration of the activity are also associated with greater amounts of inhaled $PM_{2.5}$, when compared to shorter activities undertaken along more polluted busy roads. Moreover, whereas the increased inclination of the track translates into higher inhalation rate, the amount of inhaled pollutant is still lower than when running along a major street.

The study advices that in order to minimise the individual exposure to air pollution, runners should choose shorter and faster routes. The limitations of the approach presented in the study support the need for further investigation of the long-term exposure to air pollution of urban citizens.

Word Count: 11 748

Acknowledgments

This dissertation would not have been possible without the support that I have received from Dr James Cheshire. Since the beginning of my studies, James has helped me to develop my data analysis and visualisation skills, as well as an interest in data science. His guidance, advice and comments have been essential in the process of writing this dissertation. I am very thankful for having a possibility to learn from him, which I believe will be crucial in my future career.

I would also like to thank Dr Mirco Musolesi, who has helped me in developing the ideas and methodological approaches for this dissertation.

I would also like to thank my personal tutor, Professor Jonathan Holmes, who has supported me greatly since the first day of my time at UCL. I am very grateful to him for empowering my academic development. From the very first day at UCL, Jonathan has always been keen to help me with any sort of difficulties, either personal or academic. For this I am sincerely thankful.

I would also like to thank my sisters, who took care of my well-being. Their constant support allowed me to get the most of my studies. I am especially grateful to Mary who helped me to adopt to a new environment and has been supporting me in everyday life.

I am also thankful to my girlfriend, Nana who has always been there for me. Her constant support and understanding have been invaluable. I am grateful to her for believing in me and taking care of me.

I would also like to mention my course-mates and friends from Poland, who allowed me to push my limits. I wish all of you the best and I am sure that there are still great memories ahead of us.

Finally, I would like to thank my parents for their constant and unconditional support. I would have not written this dissertation without having an opportunity to study abroad. Their encouragement and belief gave me a lot of confidence in my abilities and allowed me to develop my interests. I am very thankful to them for supporting my education since my early ages. Where I am now, is all thanks to them.

List of Contents

List of Tables

List of Figures

List of Acronyms

API	Application Programming Interface
CO	Carbon monoxide
GPS	Global Positioning System
LAEI	London Atmospheric Emissions Inventory
LAQN	London Air Quality Network
NO_2	Nitrogen dioxide
NO_x	Nitrogen oxides
O_3	Ozone
$PM_{2.5}$	Fine particles with a diameter of 2.5 micrometres (μm) or less
PM_{10}	Fine particles with a diameter between 2.5 and 10 micrometres (μm)
UCL	University College London
ULEZ	Ultra Low Emission Zone

Preface

Running is considered to be the most common physical activity that people undertake; it does not require any preparatory training or specialized equipment. The easiness of going out for a run means that there are also a wide range of motives that encourage people to run; for instance, to improve the physical condition, relax, socialize, sightsee or commute to work. The fact that running can take different forms means that it is often more complex than we think. In this sense, how and where we run might be dependent on whether we are running alone or in a group, weather conditions, traffic volume, current physical condition and well-being, mood or other often unrelated issues. These are important factors because they significantly and often imperceptibly influence our running routes. This complexity becomes more evident in urban dwellings, where all of the discussed aspects have a bigger magnitude – it is easier to find other fellow runners, the traffic volume is heavier, weather conditions are even more influential. More importantly, in times of increasing concern of sustainability and climate change, urban dwellers have become more worried about their exposure to air pollution.

As a keen amateur runner, I have spent a vast amount of time figuring out a perfect running route that would meet all of the requirements that are particularly important for me – an uncrowded running route with a stable inclination, a ground that is firm enough to allow run fast, but not too firm so my joints are not endangered, and associated low air pollution along it. Knowing that the best way to test a route by its inclination, traffic volume and built infrastructure is to examine it empirically, I also knew that it is not possible to assess the associated air pollution. Moreover, aware of the interplay between all of these features, I have concluded that it is not only important how high is the air pollution concentration along each track, but how such aspects as time, inclination, infrastructure and traffic volume affect the amounts of airborne pollutants that we actually breathe in. In this sense, while the infrastructure and inclination influence the speed with which we move and the breathing rate, the duration of the journey affects the exposure to air pollution. Having this in mind, one can possibly ask following questions: does it still make sense to run up to the top of Primrose Hill, knowing that relatively flat Prince Albert Road is associated with a lower ventilation rate? Is it worth taking a slightly longer but less polluted route, rather than much shorter but highly congested road? How long does one have to run in Regent's Park to overcome the cost of getting there by running along Euston Road? These are only examples, but they are crucial for understanding that running and mobility in general is intertwined with making deliberate choices. As a Geography student who comes from a country struggling with one of the highest air pollution concentration in Europe, it was particularly important for me to undertake a research that would provoke urban dwellers to start thinking how they can minimise their exposure to air pollution by changing their everyday movement customs.

1 Introduction

Running is one of the most popular and the most flexible type of physical activities that requires marginal or even no costs. While other disciplines attract more and more people every day, providing an interesting mean of exercise, running is still seen as a fundamental part of every physical activity. Roads, pavements, parks and paths in cities are colonised by runners of different gender, age, socio-economic status or physical condition (Latham, 2015). As the physical distance between people in urban dwellings is diminished to the minimum, the way we use the remaining free space is what defines life of cities and their residents (Gleaser, 2011; Latham 2017). Whether they do it consciously or spontaneously, runners choose their running routes having in mind different urban factors (Hitchings and Latham, 2016). For instance, Kamargianni (2015) notices that cyclists are less likely to cycle in cities when the weather conditions are bad, or the traffic volume makes it dangerous or unpleasant to ride along the main road. Similarly, runners make decisions on their running routes taking into account aspects of urban life that are important for them – for some runners it is weather, for others it is aesthetic values of the route. Nevertheless, as more and more urban citizens are concerned about the health implications of high air pollution that is present in cities, runners and also academics started to pay a greater attention to how their outdoor physical activity relates to the ambient air pollution.

The best examples of such studies include researches undertaken by Milton and Steed (2005, 2007) and Woodcock and his colleagues (2014a) who examined exposure to air pollution of pedestrians and cyclists, respectively. In spite of being key milestones in the research on the relationship between physical activity and air pollution, both studies miss out a few vital properties of running in urban areas.

In this sense, Woodcock *et al* (2014a) focused in their analysis on cyclists from the bike sharing system in London. While data on such systems are often extremely comprehensive and robust, the tracks undertaken by the cyclists were modelled and thus were hypothetical, meaning that the study constrained the users of the scheme to choose routes that are the quickest and the most optimal. As a result, the research made a strong assumption that the way in which people cycle is fixed, thus omitting leisure activates and therefore weakening the external validity of the results.

On the contrary, Milton and Steed (2005, 2007) analysed the real-time exposure to carbon monoxide (CO) by tracking mobile pollutant censors. While the study managed to capture flexibility and unpredictability of human mobility, it is difficult to apply it when studying the exposure of runners. The reason why it is so is the fact that it is simply impractical and impossible to equip runners with the mobile pollution sensors. In this sense, the research investigated the exposure to one airborne pollutant of only certain number of participants. Furthermore, considering the "increasing evidence pointing to long-term exposure to air pollutions being more important than exposure to short term peaks", it is argued that studies that use personal monitoring devices fail to capture the long-term effect of air pollution on health of urban dwellers (King's College London, 2018b). In fact, it is unrealistic and unreasonable to investigate the long-term exposures to air pollution by focusing on mobile monitoring sensors.

The research proposed in this paper aims to address all of these issues and examine how should we choose our running routes to run in the most air-friendly areas. Building on existing knowledge, the research analyses real-time data of Strava[1] users in London and their personal exposure to air pollution. Apart from focusing on the accurate real-time data of runs completed by the research participants, the research takes into account the significance of inclination and duration of an activity on the actual exposure to air pollution. While existing literature has accounted for the importance of time in the exposure to air pollution, the effect inclination is rarely recognized as an influential factor that impacts the amount of air pollutants that one actually breathes in. The knowledge on the effect of inclination on the ventilation rate is particularly important for studies that do not use personal monitoring devices, as the inhalation might differ depending on the undertaken route thus affecting the exposure to air pollution. Moreover, by using modelled data of high resolution on annual average of air pollution, the research examines the harms associated with long-term exposure

[1] Strava is one of the most favourite fitness applications in the running community (Munn, 2018). This mobile fitness app used by runners all over the world uses Global Positioning System (GPS) to track and record information on physical activity of a user, at the same time providing robust data on duration, pace, time, elevation and actual route of an activity.

to air pollution and makes it easily applicable for future researches that will elicit access to millions of records recorded with a fitness application.

The paper is divided into six sections. In Chapter 2 I will review the existing literature on the topics related to the research questions. Later I will explain an appointed methodological approach in Chapter 3 and discuss data utilised in this research in Chapter 4. Further, Chapter 5 will present the steps of data processing that have been undertaken before the process of data analysis. Finally, I will present the results in Chapter 6 and discus their relevance in Chapter 7.

2 Literature Review

As a central aspect of human mobility, running has an effect on both cities and their citizens. Activities such as running or walking are basic modes of transport in urban dwellings, shaping and animating lifestyles of the citizens in different ways (Hitchings and Latham, 2017). Regardless of lack of space or enhanced traffic volume, “urban-based adult populations are more active than regional- and rural-based adult populations” (Short *et al,* 2014: 1320). As a result, the existing literature focused principally on benefits of being physically active in cities. As suggested by Pucher and his colleagues (2010), exercising regularly for even less than an hour translates into substantial health benefits. In fact, the fact that exercising has a positive influence on one’s physical and mental health seems to be undisputed (Ludwig *et al,* 2015). Nevertheless, promoting physical activity in urban areas, for instance through a development of public cycle hire schemes, has also a positive effect on urban-related aspects, such as traffic volume or air pollution (Latham and Wood, 2015). In this sense, Winters *et al* (2007) found a positive impact of promoting physical activity in cities not only on the reduction of the risk of persistent diseases, but also on improving ambient air quality. Moreover, majority of the research on geographical aspect of running in urban areas has concentrated on the significance of infrastructure and built environment in the activity of urban dwellers. On one hand, it is suggested that the higher park density in cities, the more active citizens are (Sallis *et al,* 2016). On the other hand, Ali *et al* (2017) argue that the spatial distance to parks is in fact negatively correlated with the physical condition. Interestingly, Piatkowski and Marshall (2015) took prominently different perspective arguing that physical activity is rarely influenced by the development of urban infrastructure. Nonetheless, these are only a few examples of the exhaustive number of researches that focused on physical activity in cities. What one should bear in mind, is the fact that there is still a big gap in the knowledge on physical activity. The biggest critique of Cook and his colleagues (2016) is the fact that the existing literature focuses primarily on competitive running, omitting more popular recreational running. In this sense, while Scott *et al* (2003) looked at professional athletes analysing how natural environment of their place of origin influences their condition, other researches turned into marathon runners and their characteristics (Carter, 2015). In fact, a great attention was paid to the demographical aspect of running. Whereas Hoffman and Fogard (2012) focused on ultramarathoners, Hanson and Buckworth (2017) studied the characteristics of barefoot runners. What is worth noting, is the fact that significant majority of researches have focused on qualitative approach to the topic. As a result, apart from national and consumer surveys, there is a little focus on the statistical analysis of runners.

2.1 Physical activity and pollution

Considering the growing popularity of running and cycling as means of commuting, majority of academics have focused on how air pollution affects physical activity and its benefits, both environmental and health. In this sense, promoting physical activity in urban dwellings is seen as one of the most efficient and sustainable approaches to improve air quality and health of urban citizens (Winters *et al,* 2007; Latham and Wood, 2015). In fact, health benefits of physical activity are undisputed. However, while Pucher *et al* (2010) noted that even exercising regularly with a modest intensity translates into significant health benefits, it was also noticed that the risk of injury might be even higher and the impact of pollution on health even more severe when exercising in urban areas; in fact, physical activity in urban areas has been found to be linked "with an increased inhaled dose of fossil fuel-derived black carbon" (Nwokoro et al, 2012: 1091). As a result, it has also been suggested that air pollution often discourages dwellers to undertake outdoor physical activity (An *et al,* 2018). Although some studies found that the higher risk of injuries and more severe effect of air pollution that are associated with physical activity are in fact outweighed by health benefits, the concern of the exposure of runners to air pollution is still present (Lindsay *et al,* 2011).

The fact that high levels of air pollution negatively affect health and eagerness to exercise seems to be unquestionable. In fact, over a million of deaths in cities all over the world each year are directly caused by the exposure to high levels of air pollution (Dias and Tchepel, 2014). Moreover, as air pollution that is significantly above recommended levels affects all citizens of major cities leading to chronic diseases and lung cancers, thousands of British citizens die prematurely and even more children are in danger of the neurodevelopment complications (Lewis and Edwards, 2016; Buechler, 2018; Stingone *et al,* 2017). The reason why air pollution is so dangerous for people is the fact that inhaled pollutants severely affect human respiratory system, intoxicating the body and reducing the circulation of oxygen in blood (Mead *et al,* 2013). The pollutants that are especially harmful include such chemicals as NO_2, CO_2, O_3, PM_{10} or $PM_{2.5}$. In spite the ongoing debate on which one of them is the most dangerous, most of the scientists agree that the latest of them is associated with the most adverse impact (Gulliver and Briggs, 2005). In fact, $PM_{2.5}$ is "one of the most abundant air pollutants" that "could penetrate the deepest part of the lungs bronchioles or alveoli and result in asthma, lung cancer, respiratory diseases, cardiovascular disease", meaning that it became defined as a pollutant that is worth the biggest concern (Zhao *et al,* 2018: 1; Zhu *et al,* 2018). Consequently, long-term exposure to $PM_{2.5}$ is assumed to link to 8% of global mortality (Mila *et al,* 2018).

2.2 Use of GPS technology

The development of the GPS technology resulted in the flourishment of studies that focus on human mobility. In this sense, GPS records are commonly used to investigate the trajectories, predict traffic volume or calculate gas emissions of various modes of transport – cars, public transport or bicycles (Evans *et al*, 2014; Lipar *et al,* 2016). Moreover, the technology has been more commonly used in studies concerning human bodies and their responses to different environmental factors. In this sense, GPS records have also been used to empirically test variability in heart rate or exposure to air pollution during outdoor physical activity (Hunt and Hunt, 2016). Nevertheless, the research has taken two approaches to calculate short and long-term personal exposures to air in the context of mobility. On one hand, some studies have analysed GPS data combined with personal monitoring devices, to accurately capture the variability of the exposure throughout the activity (Milton and Steed, 2007; Lu and Fang, 2015). On the other hand, some of the academics have studied the exposure to air pollution across the city by modelling the most likely routes based on the exploitation data of the public bicycle hire schemes (Woodcock *et al,* 2014a). More importantly, existing literature has done a little in combining the two approaches by using GPS trajectories in relation to levels of air pollution; such methodology is particularly desired considering the evidence for the significance of long-term exposure to air pollution.

2.3 Ventilation and inhalation

Nonetheless, while the needlessness of the personal monitoring devices is seen as a great advantage, it also poses a key problem of lack of information on the actual intake of airborne pollutants. In this sense, Woodcock *et al* (2014: 4a) found that the average air pollution concentrations for walking and cycling were lower than for commuting by public transport, but these were "more than counterbalanced by higher ventilation rates". Considering that both high heart rate and inhalation of physically-active commuters are related to higher amount of inhaled pollutants, there has been a growing focus on the role of kinetics in exposure analysis (Cepeda *et al,* 2017; Barstwo and Mole, 1991). In this sense, many studies showed the that by modelling the intensity of the exercise and related variability of heart rate, one can effectively estimate the oxygen consumption, a fundamental aspect of the air pollution exposure (see Ludwig *et al,* 2015; Stirling *et al,* 2005; De Brabandere *et al,* 2018). In fact, while Kim *et al* (2018) argue that peak of the oxygen consumption is a key characteristic of exercise physiology, Baty *et al* (2016) and Haddad *et al* (2014) correspondingly claim that the relationship between heart rate and ventilation is very strong and can be accurately demonstrated by fitting using a nonlinear regression modelling. In this sense, Bergstrom *et al* (2015) and Svenby (2015) tried to capture the changes in kinetics of cyclists and both found parallel trends in the dynamics of heart rate and oxygen consumption. More importantly, other studies have found that these trends have different trajectories for different types of activities (Millet *et al,* 2009). Whereas Bonsignore *et al* (1998) looked at the breathing rate of triathletes and found the ventilation to be higher for cycling than for running, Berry *et al* (1996) concluded that running is characterised by higher ventilation when compared to walking, 73.7 litre/min versus 68.6 litre/min respectively.

One of the simplest ways to increase ventilation is to intensify the exercise, which can be easily done by increasing the inclination of the exercise; as the slope of the run increases, so does heart rate, blood lactate and ventilation (Zakynthinaki, 2015). Huck *et al* (2017: 122) explains that the reason why it does matter is the fact that exposure to air pollution will be different across individuals "whose respiratory patterns will react in different ways to journey duration and changes in elevation". Interestingly, academics have focused on the relationship between inclination and oxygen consumption for horses; studies of Schroter and Marlin

(2002) and Wickler *et al* (2000) suggested that ventilation rates of horses increase while running uphill. As Eaton *et al* (1995: 956) argued, changes in the "treadmill incline provide a reliable means of increasing the intensity of exercise without increasing the speed or altering the relationships between HR, plasma lactate concentration and VO_2". Notably, even more studies have analysed the effect of changes in inclination on the mechanical work of human body and the variabilities of heart rate and oxygen consumption when running (see Minetti *et al*, 1994). Importantly, Wallaart (1997: 4) argued that "as work done increase, the amount of air used increase" pointing to "a close relationship between the external power which is produced and the oxygen consumption of a subject". However, Padulo and his colleagues (2012: 1332) noted that "relatively few studies can be found in literature investigating the effect of uphill and downhill running from the kinematic point of view". As they argued, this is important as the biomechanical aspects of running (such as a length of step or a step frequency) have significant influence on the ventilation rates. In fact, they argue that uphill running is characterised by higher exhaustion, increased step frequency and therefore higher ventilation (Padulo *et al,* 2013). Importantly, the authors have found a positive relationship between inclination and ventilation rate, suggesting that when compared to running at an even ground, running at the 2% and 7% inclines is associated with increase in oxygen consumption of 10% and 19% respectively (Padulo *et al,* 2013).

Finally, the literature recognizes the importance of the exhaustion and increased ventilation rates on the actual amounts of airborne pollutants inhaled by urban citizens. While Lu and Fang (2015) see inhalation as a key channel through which airborne pollutants penetrate human respiratory systems, De Nazelle *et al* (2013: 93) note that the estimation of "inhaled dose of air pollution by accounting for energy expenditure can change the relative ranking of exposures as a function of activity patterns compared to using solely exposure concentrations".

2.4 Temporal aspect of exposure to air pollution

Nonetheless, the literature does also identify that "exposure is rarely constant" (Brauer, 2010: 111). As Lu and Fang (2015) observe, exposure to air pollution ultimately happens at a time. This means that it is essential to understand that one's exposure is primarily dependent on the duration of the exposure to air pollution. In this sense, Gulliver and Briggs (2005) see that the time of the activity is in fact important, but partly unstudied element of exposure to air pollution. Liu *et al* (2013) noticed that activity patterns play a great role in the duration of each physical activity, meaning that it is important for the actual exposure to air pollution where and for how long runner stops. Nevertheless, while De Nazelle *et al* (2013) argue that there might be substantial differences in exposure to air pollution throughout the day, they also maintain that the existing literature rarely captures all of the components of mobility that should be taken into account; accordingly, Mila *et al* (2018) found that the human exposure can differ depending on one's gender or location of the activity. Considering temporal dimension of the exposure rates, the academics have recently focused on the development of a methodology that would account for both spatial and temporal aspect of exposure to airborne pollutants; in this sense, Gerharz *et al* (2013) and Zakynthinaki (2015) proposed models that justify the role of time and nature of the physical activity in modelling human exposure. More importantly, Dias and Tchepel (2018: 2) has proposed to quantify exposure "as a function of concentration and time". In their recent research, the authors argue that there is a need to use a time-integrated exposure rates that would allow to accurately calculate exposure rates based on the duration of the activity.

3 Methodology

In order to understand the nature of the studied runs undertaken by the participants of the research, one should firstly examine where these runs did actually take place. By exploring the traffic volume of the participants, one can easily spot where did the participants usually run or what were their most favourite routes. These are important questions as the spatial patterns can often discover the characteristics of the runners and tell us a little bit about their habitats and behaviours.

3.1 Heatmap technique

To capture the spatial patterns of physical activity, a technique of heatmap is recommended. This method of geographic data visualization uses different shades of a colour to show the variations in the intensity of the studied phenomena. In this sense, the areas that are characterized by activities of higher frequency are represent by darker shades, while the areas less visited less often have lighter shading. The technique of a heatmap is also appreciated by Strava, who has created a global map for almost all activities recorded with their mobile application (Figure 1).

Figure 1. Heatmap of activities recorded by Strava users. Source: Mapping London (n.d.).

The map clearly points two great advantages of this approach. Firstly, the heatmap captures all activities undertaken all over the world, meaning that the tracks are not spatially limited. Secondly and most importantly, it is clear that this method visibly demonstrates the spatial patterns of physical activity.

3.2 Exposure to air pollution

A common approach to investigate one's exposure to air pollution is to analyse the concentration of air pollution along the route of interest. When looking at the pollutant concentration, two different routes can be compared in regard to the air quality by simply matching the pollution levels along each of the track. While such approach might be useful when examining the exposure to air pollution of pedestrians, it seems too primitive and general for the investigation of the exposure to air pollution of runners. The reason why this might be the case is the fact that it is not the concentration of pollutants in air that matters, but rather one's inhalation rate and the amount of airborne pollutants that one actually breathes over time (Buechler, 2018). As Dias and Tchepel (2018: 3) argue, the exposure to air pollution and the actual concentration are distinctive things which differ by the human factor, often meaning that the "high air pollution concentrations do not necessarily result in high exposure".

In this sense, pedestrians and runners do differ not only by their heart rate and ventilation, but also the way how the gradient affects their exposure to air pollution. In fact, the heart rate and ventilation of pedestrians are unlikely to change substantially with an increase of gradient, as they can easily adjust their tempo and effort to the conditions of the environment; this means that ventilation of pedestrians is thought to be constant across time. Contrary, with an intent of improving their conditions, runners do not slow down when they know that they might get tired. This also means that changes in the elevation profile have a substantial effect on runners' ventilation. In other words, the slope of a hill does matter for runners as it may lead to an increase in their ventilation and therefore the amount of inhaled pollutants. This is also the reason why runners' ventilation does matter: "the faster you breathe the more airborne pollutants are delivered to your lungs" (King's College London, 2018c). Moreover, this suggests that the routes that are associated with an increased ventilation (for instance uphill tracks) can also be more hazardous for our respiratory systems, regardless the concentration of pollutants in the ambient air is low or high. Having this in mind, instead of just looking at the raw data on the ambient air pollution along the route, the research will turn to predicted amounts of pollutants inhaled by runners.

To account for the importance of a ventilation rate in the actual exposure to air pollution, the research will make use of the work of Woodcock and his colleagues (2014a: 3) who "estimated the exposure to $PM_{2.5}$ along each route, by applying published estimated of average 24-hour $PM_{2.5}$ concentrations in a $20m^2$ grid across central London." The equation proposed by the authors consists of 4 variables and can be notated as follows:

$$\textit{Exposure rate} = \textit{average } PM_{2.5} \textit{ concertation along route} \times \textit{ventilation rate} \times \textit{road positioning scaling factor} \times \textit{pollution composition factor}$$

where *ventilation rate* is the amount of air inhaled by a commuter, *road positioning scaling factor denotes* the position of an individual on the road and *pollution composition factor* informs whether the measurement was taken on the surface, on in the underground (Woodcock *et al,* 2014b).

Nevertheless, the equation suggested by the authors can be simplified considering the fact that the analysis focuses solely on the activities that were completed by runners on the surface; this means that the factors of *road positioning scaling* and *pollution composition* can be ignored as they are not relevant for this research. Moreover, as discussed before, the focus on the mean value of pollutant concentration is also subject to questioning; although the average $PM_{2.5}$ concentration along route might give some idea of the quality of the route, it is still quite imprecise and uninformative. By simply looking on the average concentration along route, we omit much of the variation of air pollutants. Therefore, instead of using the average pollutant concentration, the equation will use an actual value of air pollution for a given location. In this sense, the revised exposure rate proposed by Woodcock and his colleagues (2014a) does account for the importance of place in the exposure to air pollution and can be notated as:

$$\textit{Exposure rate} = \textit{pollutant concentration} \times \textit{ventilation rate}$$

3.3 Ventilation rate

The variable of the ventilation rate is particularly important in this research as it reflects the influence of changes in the elevation profile. While Woodcock *et al* (2014a) used the variable to account for a difference between pedestrians and cyclists (either value of 3.3 or 6.8), in this research it will capture the changes in the intensity of an activity caused by running uphill. To capture these changes, a very basic model that predicts the ventilation rate given the slope will be introduced. The model draws on the work of Padulo and his colleagues (2013) who investigated the biomechanical aspects of running and analysed how the heart rate and oxygen consumption change depending on the gradient. In their study, the authors have investigated the runs of both marathon and amateur runners and compared their results for 0%, 2% and 7% slopes. What they concluded was that when running at a 2% incline, the ventilation rate and oxygen consumption increase by 10% compared to running on an even ground. Similarly, they have also agreed that running at a 7% incline is associated with an increase of 19% in the ventilation rate and oxygen consumption, compared to running at a 0% incline.

The model based on the findings of Padulo *et al* (2013) was transferred into a following equation:

$$f(s) = \begin{cases} 1 + 0.05 \times s, & s < 2 \\ 1.068 + 0.016 \times s, & 2 \leq s \leq 52 \\ 2, & s > 52 \end{cases}$$

where *f(s)* is ventilation rate, *s* is slope and s is nonnegative. In this sense, an increase of one unit in slope is associated with a 0.05 increase in the ventilation up to 2% of slope, and then an increase of 0.018 in ventilation for every unit increase in slope. Moreover, the equation accepts that the increase in ventilation rate cannot be higher than 2; it is assumed that average heart rate when walking is half of the maximum heart rate and as the ventilation rate and heart rate are closely correlated thus it can be concluded that the ventilation rate cannot be twice higher than the average. As a result, for every point with a slope higher than 52%, the ventilation rate is equal to 2. For the purpose of this research, it is assumed that the relationship between ventilation and gradient is linearly interpolated (see figure 2).

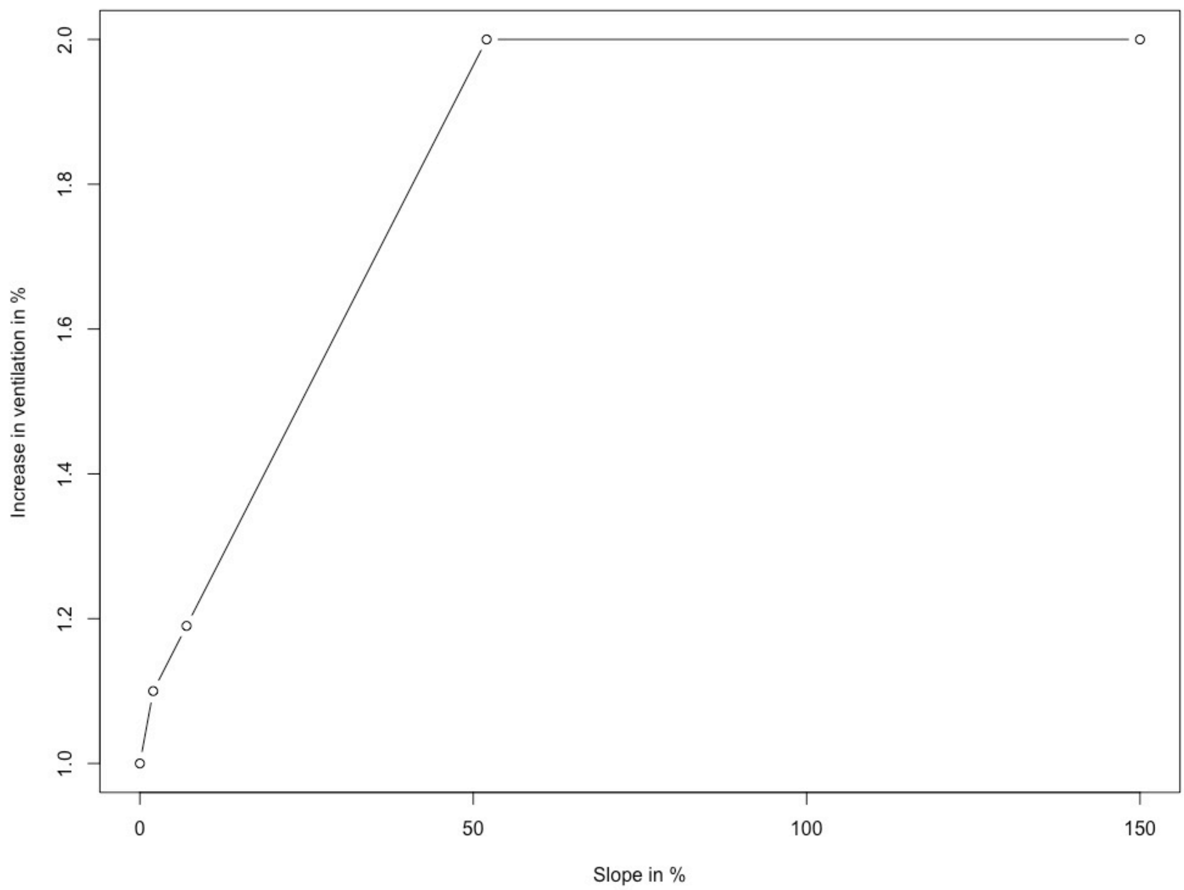

Figure 2. Graph of the linear interpolation of a relationship between ventilation rate and slope.

The reason why the relationship is calculated with a linear interpolation method is the fact that it is unknown how exactly the rate of an increase in ventilation would change when slope is higher than 7%. It is suspected that the rate of change in the ventilation rate would decrease translating into lower effect of the slope on the ventilation rate. On the other hand, it is possible that extreme values of a gradient would lead to a very high ventilation rate. Although it is unlikely, I will adopt a more pessimistic approach where the relationship between ventilation rate and slope is linear when slope is greater than 2% (this should better emphasize the areas with high ventilation rates). Nevertheless, one should bear in mind that such approach is problematic as it does not explain how ventilation actually increases with an increase of slope.

3.3.1 Ventilation rate modelling

What needs to be underlined is the fact the suggested model for a ventilation rate makes two strong assumptions that substantially threaten the external and internal validity of the results. First and foremost, it assumes that runners move with a constant tempo and their effort does not change over time, meaning that their heart and ventilation rates are not affected by the duration of an activity. This assumption is more than problematic because in reality one's exhaustion increases with time; it is axiomatic that the longer we run, the lower the intensity of an activity is (Stirling *et al,* 2005). This also means that it is impossible to maintain the ventilation rate constant over time.

Secondly, the model assumes the complete homogeneity across the population meaning that the predicted changes in the ventilation rate are independent from the personal characteristics of a runner. In reality, some runners can be healthier and have better physical condition than other runners. Moreover, the "respiratory patterns" of some runners "will react in different ways to journey duration and changes in elevation and meteorological condition" (Huck *et al,* 2017: 121).

Nonetheless, the proposed model has also a big advantage: it successfully solves the problem of lack of data on inhalation of runners. In fact, it is impossible to record the amount of inhaled air by runners who undertook their runs in the past. Moreover, although informative and reliable, I argue that measurements of respiratory changes are not practical as they allow to study only those who run with devices designed to take such measurements; in real life, running with such device might be inconvenient and significantly change the efficiency of a runner. However, as we assume that the effect of a ventilation rate on exposure rate is constant across the population, it actually does not matter how much airborne pollutants we breathe in, but rather the difference between the ventilation rates. In other words, all one needs to know is how the ventilation rate changes with an increase in a gradient; whether it is 20, 35 or 50 cubic feet per minute, the rate of an increase in the ventilation rate will be the same for every individual. As a result, we do not have to possess data on actual values of inhalation, which means that we can easily imply the exposure rate model to every track for which we have information on the ambient air pollution.

3.4 Time and exposure

As discussed in Chapter 2, time plays a great role in both physical activity and the analysis of exposure to air pollution. To really understand exposure of runners to air pollution, it is essential to examine temporal aspect of the exposure (Gulliver and Briggs, 2005). In this sense, it is also important for how long one is actually exposed to air pollution as the longer one is exposed to air pollution, the more airborne pollutants one inhales. To go further, Dias and Tchepel (2018: 559) see exposure "as a function of concentration and time" and propose time-integrated approach which understands the exposure as "the integral of instantaneous exposures[2] over the duration of exposure", expressed by the equation:

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

> where E_i is the time-integrated exposure experienced by the individual i, C_i (x, y, z, t) is the concentration occurring at a particular point occupied by the individual i at time t and spatial coordinates (x, y, z) corresponding t_1 and t_2 to the starting and ending times of the exposure event, respectively (Dias and Tchepel, 2018: 559-560).

In order to calculate a numerical approximation of the time-integrated exposure for the chosen timeframe, a chained trapezoidal rule of a numerical analysis has been used. In this method, the exposure is calculated by dividing the time interval and then summarizing the results of individual segments. Each of segments is represented as a trapezoid, where the pair of bases are the pollution intake values and the height are the difference in time between these two points (see figure 3). The numerical approximation of a segment is calculated as a surface area of the constructed trapezoid [$S = \frac{1}{2}\ (a + b) \times h$] and after the transformation of the equation for a trapezoid's surface, can be written as:

$$S = \frac{1}{2}\ (p_{i+1} + p_i) \times\ (t_{i+1} - t_i)$$

[2] At the same time, Dias and Tchepel (2018: 559) define instantaneous exposure as "the exposure at an instant in time" which is also "expressed in the same unit as the concentration (e.g., µg·m^{-3})".

where S is a surface area of a trapezoid, p_i is the concentration at point i, and t_i is the timestamp at point i. By summarizing the area of trapezoids, one can get an approximate value of the area surface of the interval of interest.

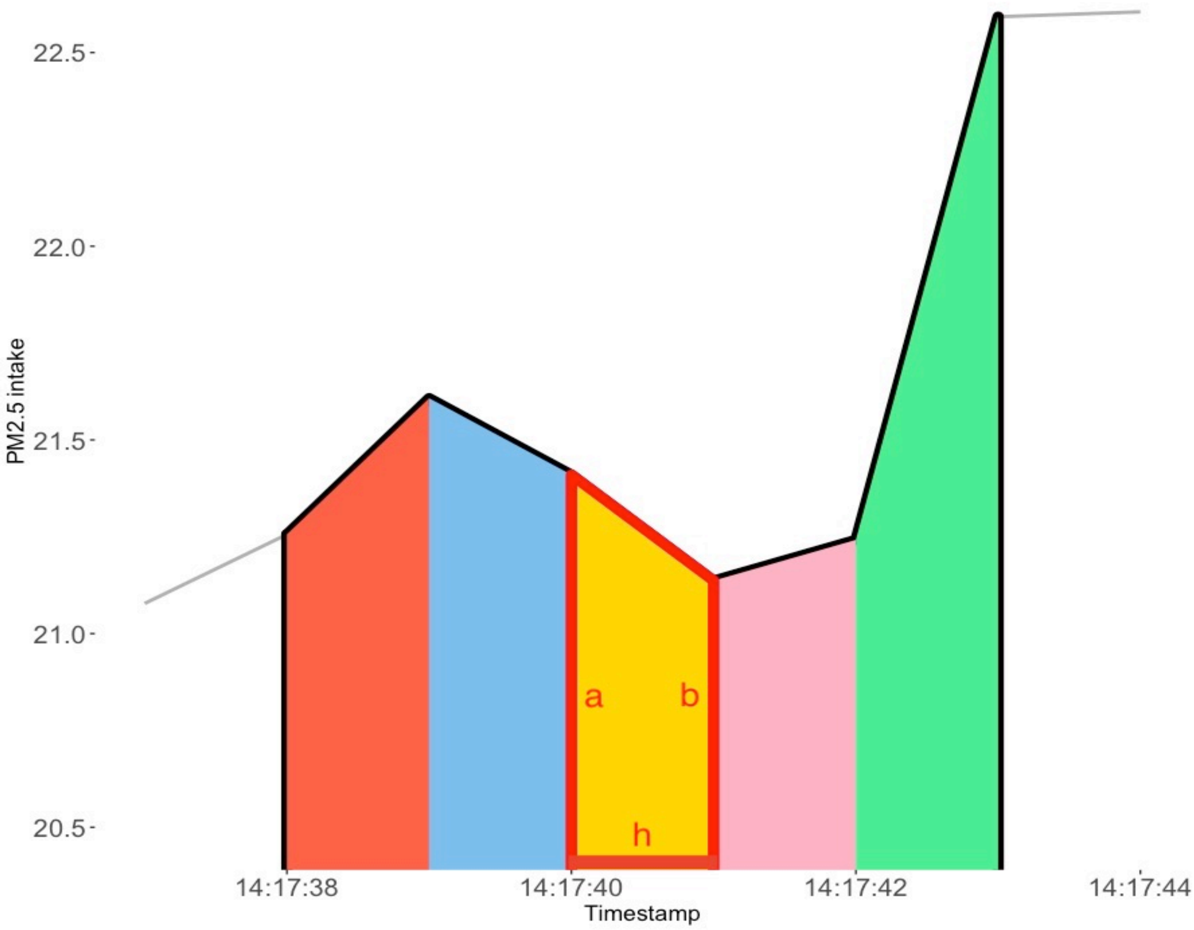

Figure 3. Graph of an example of the chained trapezoidal rule.

4 Data

The study focuses primarily on runs that have been completed with Strava. With a development of wireless mobile phone generations (particularly 3G and 4G), fitness apps became more popular among the runners as they gave them opportunity to better understand their mobility (Wright *et al*, 2008). Nowadays, running with devices such as mobile phones, smart watches or GPS running monitors that log one's physical activities is so common among amateur and professional athletes, that over 80% of runners use mobile apps to gather information on their activities (Running USA, 2017). As a result, professional athletes record up to 224 days of training on average, while amateurs log an average of 85 days per year (5kevents.org, n.d.)

For this research I have used data from five volunteer Strava users. All of the participants have completed at least one run in London and are affiliated with University College London. Due to a development of a global heatmap that have exposed secret military bases, there are ongoing concerns about the security and privacy of Strava users (see Hern, 2018). Having this in mind, each of the participant was given an informed consent form that regulates how their data would be used in the study. In this sense, none of the acquired data contain sensitive personal information (Strava Metro, 2015). Furthermore, none of the participants disagreed to use their data in the visualization of the results. To anonymise the tracks, all of the records were aggregated. After excluding activities undertaken outside of London and those saved in a wrong format, the research analyses 195 records.

4.1 Air pollution data

Essential for this study is robust data on air pollution. Fortunately, London provides one of the most reliable, robust and up-to-date datasets on various pollutants. As a part of the programme to raise awareness about the air pollution among its citizens, the city communicates current air quality by using maps that visualize variations in Air Quality Index (Lu and Fang, 2015). Although the index is easily understandable by the public, it is very limited in terms of the understanding of the ambient air pollution for specific areas. In order to study air pollution for every location visited by the volunteers, data on air pollution in London were obtained from the London Atmospheric Emissions Inventory (LAEI), a schedule that includes comprehensive data on key pollutants ($PM_{2.5}$, PM_{10}, NO_x, NO_2) within the M25 motorway. The biggest assets of the data facilitated by the inventory is an extremely high spatial resolution of data. In this sense, LAEI provides high resolution (20x20m grid cell) datasets on the modelled annual averages of the pollutants for 2013, as well as modelled averages for 2020, 2025 and 2030 (figure 3). The most recent dataset (published in 2016 and updated later in April 2017) is available to download from the London Datastore Portal and was used for this research.

4.1.1 Air quality monitoring networks

Although datasets provided by LAEI give very broad understanding of the average air quality in London, temporal resolution of the data is extremely low in a sense that temporal variations in the air pollution concentrations are not captured. To investigate daily or monthly changes in air quality in London, one can turn to air quality monitoring station networks, such as London Air Quality Network (LAQN), which is a network managed by the Environmental Research Group at King's College London that takes independent and systematic measurements of the air quality in London (King's College London, 2018a). Although the data gathered by the monitoring stations can inform about even horal changes to the pollutants' concentration, the measurements are not regular and consistent in terms of the spatial coverage and the measured pollutants. In this sense, only a few monitoring sites would have long-term and complete data on each of the pollutant of interest, meaning that real-time data on air pollution would not be helpful.

Most importantly, academics start to recognize that "exposure to short term peaks" of air pollution is actually less influential and harmful than "long-term exposure to air pollution" (King's College London, 2018b). In fact, considering the adverse influence of such pollutants as $PM_{2.5}$ or NO_2 on human respiratory system, it is argued that long-term exposure is much more important and therefore the annual averages of the air pollution "are more adequate than short-term exposure" in the investigation of the influence of the air pollution on health (Mead *et al,* 2013: 195). It is also worth noting that the complexity and robustness of discussed datasets are in fact the strongest merits of the research in a sense that they have very high precision and spatial resolution.

5. Data processing and analysis

Due to the variation in the precision of the GPS technology, data has been thoroughly processed. Standard GPS transmitters, such as mobile phones, can calculate the location with an accuracy of a few meters (Karsky, 2004). Nevertheless, Yin and Wolfson (2004) noted that in reality these calculations are very often imprecise meaning that documented locations are often different from those actually visited. In fact, a data point recorded by GPS devices can have a location error that ranges from a few to hundreds of meters (Yin and Wolfson, 2004). The reason behind this is that the GPS technology needs at least 4 satellites for precise measurements; as the number of satellites whose signal reaches the GPS device decreases, so does the accuracy of the measurement (Van Dijk and De Jong, 2017). As Milton and Steed (2007) explain, the connection between the device and the satellites is unstable because the signals coming from the satellites are often altered by skyscrapers or urban canyons. In this sense, the clarity of the sky plays a great role in the accuracy of a GPS receiver, especially in urban areas that weather conditions are often fluctuant (Milton and Steed, 2007).

5.1 Data correction

Having all of these in mind, it is fundamental to take the unreliability and fallibility of GPS technology into account. In regard to this research, it is crucial to effectively treat the measurement error so that the results are more reliable and accurate. As Robb (2017) notes, most of the mobile devices, particularly smartphones, deal with this kind of problem by correcting "the GPS signal in urban areas by snapping it to know read geometry rather than a raw position". By doing this, a device assumes that an object tends to move along the fixed routes and the swerves from the predominantly known paths are results of the measurement errors. In fact, the technique of snapping points to lines is a powerful, commonly used and often seen as essential approach to data processing (Yin and Wolfson, 2004).

5.1.1 Snapping to line

Aware of this, Milton and Steed (2007) used the Ordnance Survey data to correct the errors by snapping the reported GPS positions to the closest pavement edge. As a threshold, the authors set a 20 meters radius which means that all the points within that radius are moved to the edge; in case of a distance between the point and the nearest pavement edge being higher than 20 meters, the data point is not moved.

This study utilised data from the OpenStreetMap, an open-source and free software that offers mapping data; the data were accessed using an *osmdata* package built in R. For the snapping radius a rigorous threshold of 10 meters was set. Although such a small radius means that the probability of the data point being snapped to the line is lower, there are a few reasons behind that. Firstly, the GPS technology has significantly developed since Milton and Steed (2005) published their paper. In this sense, it is believed that the development of the GPS technology means that it became more precise. Secondly, as the traffic volume plays a great role in maintaining a constant pace, runners tend to avoid moving along major roads, choosing less crowded pavements or footpaths. A relatively low threshold means that the data point has a lower probability of being snapped to a main road, while a run did actually take place on a nearby pavement. Nonetheless, even pavements and footpaths can be overcrowded. Therefore, it is argued that runners are exceptional in terms of the mobility as they often tend to avoid all sorts of established footpaths or built infrastructure. In this sense, the 10 meters radius allows a runner to move on off-road tracks (such as alleys in forests or parks). This is particularly important for runners as they often choose such tracks; it is understood that while 80% of runners run on the roads, over half of them frequent local parks or off-road tracks (Cook *et al,* 2016). The reason behind this is health issue: off-road tracks are characterized with "a more gentle and elastic cushion" meaning that the burden on knees is lower and so is the risk of an injury (Cook *et al,* 2016: 750). Another outcome of this is a change of the characteristic of a run. Although running on soften ground is healthier for our joints, it actually does take a little more effort to run on such surface; this is important as it also might affects the heart and the ventilation rates of a runner. Overall, the 10 meters radius is argued to be reasonable threshold for the process of snapping points to nearest lines as they allow for segments to be completed beyond roads and streets (see figure 4).

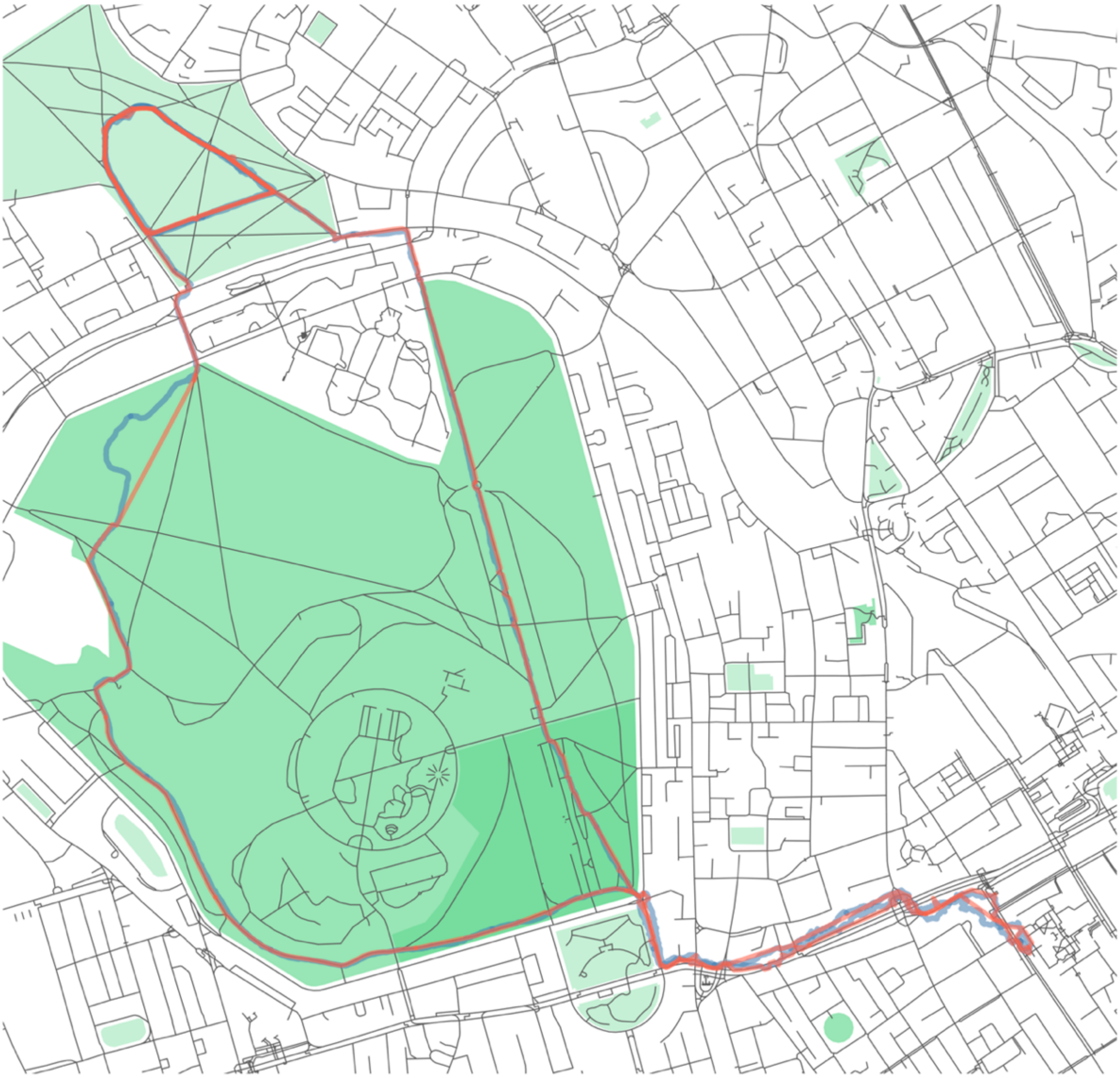

Figure 4. Map showing the difference between paths constructed based on the raw data points (blue line) and routes created with only data points that have been snapped to the roads (red line).

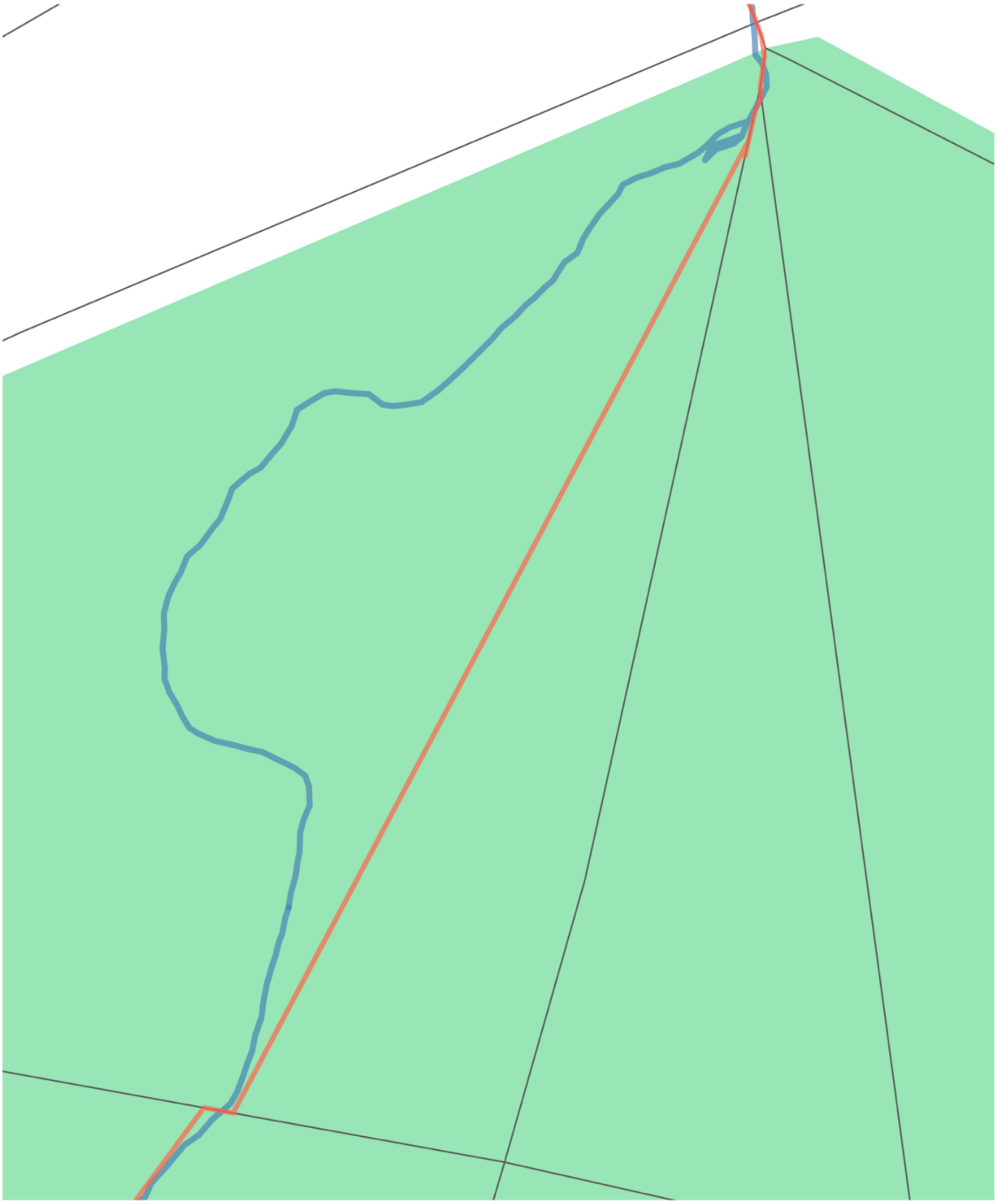

Figure 5. Map showing the difference between blue and red routes.

As the map shows, neither of these paths can accurately capture an actual route undertaken. While the blue path lets a runner free and allows them to move on unrecognized paths but also through buildings, the red line simply ignores the points that were not snapped to the route. The difference between these two paths are evident in figure 5, where the movement took place on unestablished pathway in Regent's Park. To solve the problem of such discrepancy and discontinuity of the course, two of the routes were combined; in this sense points that have not been snapped to the line were remained.

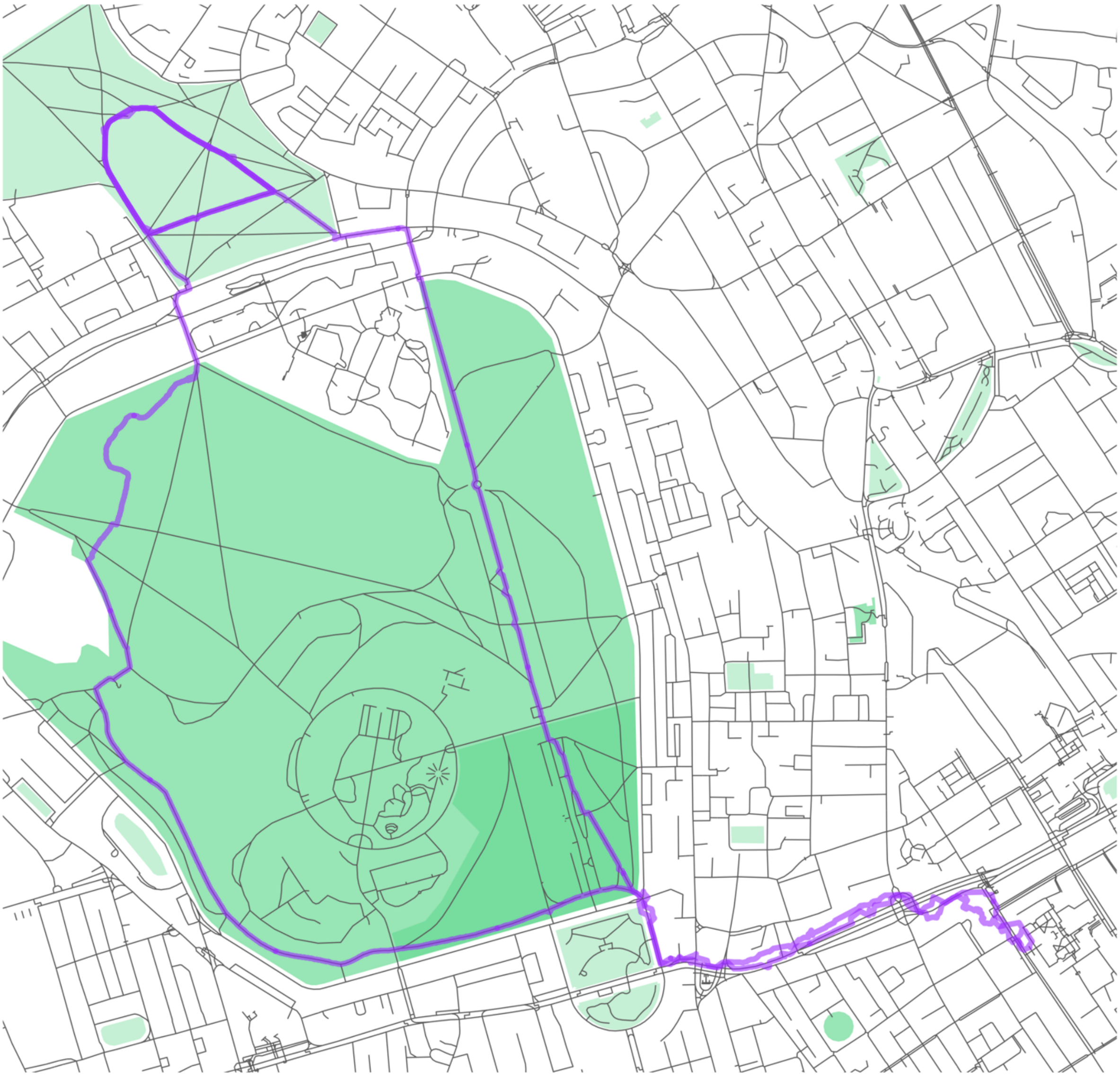

Figure 6. Map showing the route constructed by combining raw and corrected data points based on the 10 meters radius.

5.2 Activity output

As the accuracy of GPS has been corrected, it is helpful to get as much information from the records as possible. The fact that each of data points have an accurate timestamp and hold information on location, it is possible to calculate both the pace of an activity and changes in a slope along the run (De Smet *et al,* 2016); these features will be particularly useful for the detection of outliers and the ventilation rate modelling. Because the tracks were completed at different locations and time, each of the tracks was processed separately. This means that when calculating speed and slope, points from two different tracks did not end up being connected.

After the calculation, it is apparent that some values of speed or slope are suspiciously extreme. In this sense, in some cases the real-time speed is even higher than 50 km/h. Such extreme measurements are suspicious (as the world record for a footspeed is equal to 44.72 km/h and was achieved by Usain Bolt, the fastest human on earth, it is reasonable to assume that none of the participants could run faster than 50 km/h) and suggest that data may be a subject to huge errors (Guinness World Records, 2019). In fact, the following figures (figures 7A and 7B) show the suspected location errors. Figure 7B indicates that these points were not recorded intentionally by the runner but are in fact deviations from that have been logged during an activity.

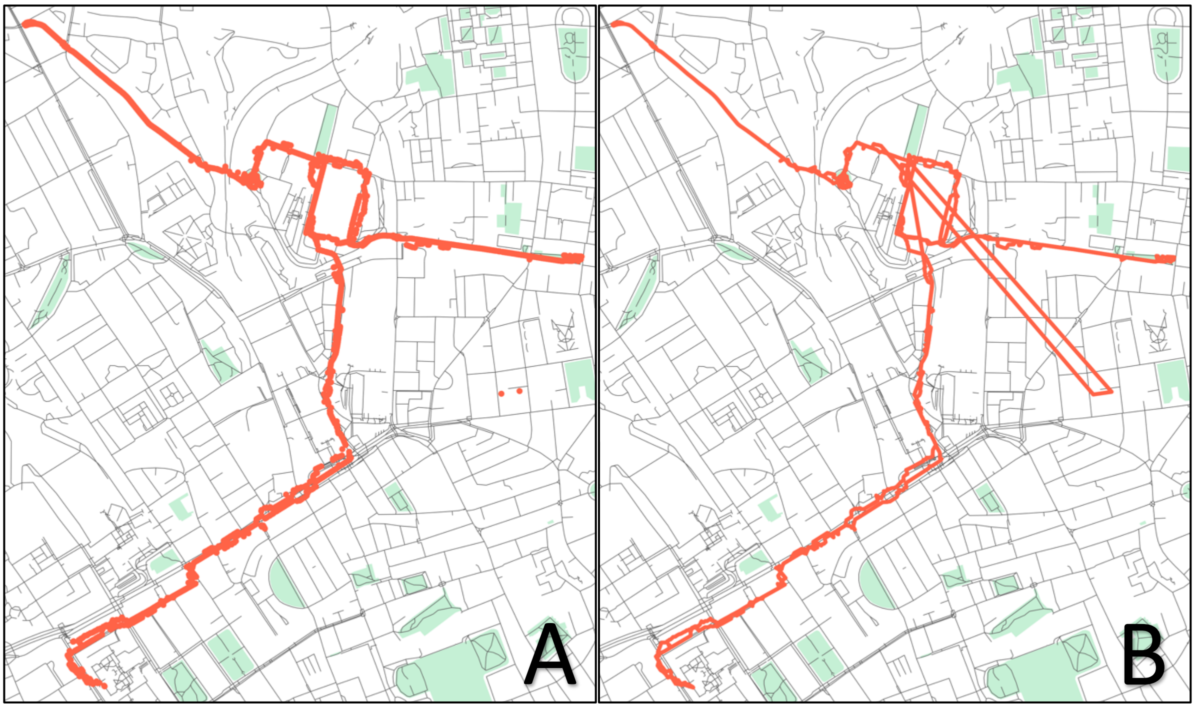

Figure 7. Maps showing the locational outliers.

5.3 Identification of outliers

To resolve this issue, an outlier was defined as a point that meets at least one of the following assumptions:

- Speed higher than 30 km/h[3];
- Spatial distance higher than 50 meters[4];
- Spatial distance lower than 1 meter[5];
- Slope higher than 150%[6].

The detection process has identified over 40,000 outliers (10.9% of all logged data points) that has been removed. Nevertheless, such harsh conditions are seen as necessary particularly in studies that rely on the precision of the data points.

5.4 Elevation modification

Nonetheless, the removal of outliers did not solve all of the data-related problems. The problem with data on elevation incorporated in the activities records is the imprecision of the measurements; the minimum value of a change in the elevation is 10 centimetres. This means that the changes in the ventilation rate may not be accurately reflected. Moreover, since the data points that were moved to roads have also changed their location, the elevation data is no longer relevant; as most of the data points got new geocoordinates, it does not make sense to investigate the elevation for coordinates that are no longer valid.

To obtain more accurate and precise data on elevation, the elevation data provided by Google were used. By using the Elevation API facilitated by Google, each of the data points has been given a new value, that is far more precise. In fact, comparing elevation data from two

[3] As argued before, it is impossible for humans to run faster than 50 km/h. Knowing that there were no professional runners in the sample, the analysis assumed that none of the research participants could move faster than 30 km/h.
[4]One way in which Gerharz *et al* (2013) defined outliers is the 100 meters distance. Nevertheless, I argue that such threshold might be inadequate considering the development of the GPS technology.
[5] Such assumption disregards points that reflect lack of movement. This is important as it directly links with the extreme slope values.
[6] This condition deals with subsequent data points that were logged on different surfaces (for instance tunnel or bridge).

different sources suggests that the slope calculated using the elevation from Google has lower standard deviation (3.76) than the standard deviation of slope calculated using the elevation obtained by Strava (5.46). This is particularly important data on slope to be as precise as possible as the ventilation model is determined by the reliability of these data.

However, as the ventilation model predicts the ventilation rate for nonnegative slope, it is meaningless to apply the model to the negative values of slopes; the effect of the negative slope on the ventilation rate is different than for positive values, simply because when running downhill we exhaust less. In fact, running down the hill does not mean that our ventilation will fall below certain level.

On the other hand, the concept of a hill is twofold – as one person may be running up to the top, the other person can at the same time run down from the top of it. In this sense, it does not necessarily make sense to investigate how one's inhalation of air pollutants change when running down- and up-hill.

Considering these two ideas, the research assumes every hill to have a positive gradient. This slightly pessimistic approach means that each slope is an ascent so that the ventilation rate is increased. To capture this, absolute values of the slope will be investigated.

5.4.1 Smoothing values

However, as the variations in slope and speed can be very high, it is helpful to smooth the results so that the changes are more transparent and understandable. To smooth the obtained results, a LOWESS technique was used, in which

> the fitted value at x_k is the value of a line fit to the data using weighted least squares where the weight for (x_i,y_i) is large if x_i is close to x_k and small if x_i is not close to x_k (Cleveland, 1981: 54).

By smoothing the values of the slope and speed, the general patterns are much easier to detect, and the variations are significantly reduced; in fact, the standard deviation of the smoothed results of the absolute values of slope is a little bit lower and equals 2.40.

To extract data on air pollution for every data point, each of the data point has been assigned the value of the raster cell within which the point lies. In this sense, every data point holds information on the concentration of $PM_{2.5}$, PM_{10}, NO_x, NO_2 together with actual inhaled amounts of these pollutants.

With clean and processed data, the creation of heatmap does not require problem-solving approach. As suggested by the Strava developers, activities have been treated as “pixel perfect paths connecting subsequent GPS point” (Robb, 2017). By doing this, activities that have more data points due to lower speed are not favourited which happens when they are treated as one-pixel records.

6 Results

Prior to the exposure analysis, the focus of the study will be defined. The analysis of the relationships between pollutants suggests a nearly perfect correlation – the correlation coefficients are not lower than 0.96 (see table 1).

	$PM_{2.5}$	PM_{10}	NO_x	NO_2
$PM_{2.5}$	1	0.98538*	0.95699*	0.97452*
PM_{10}	0.98538*	1	0.98550*	0.99256*
NO_x	0.95699*	0.98550*	1	0.99370*
NO_2	0.97452*	0.99256*	0.99370*	1

Table 1. Correlation between $PM_{2.5}$, PM_{10}, NO_x, NO_2. * p-value < 0.01.

Importantly, all of the coefficients are statistically significant at the p-value level at 0.01. Nevertheless, considering an extremely high correlation between the pollutants and the evidence for $PM_{2.5}$ to have the most adverse effect on human health, the analysis will focus on this pollutant.

Figure 8 presents the concentration map for the $PM_{2.5}$ pollutant in the neighbourhoods visited by the participants. The map shows that the highest level of air pollution in the centre of the city, particularly along the busy roads and relatively lower air pollution in the outskirts of the town; in other words, the further from the centre, the better.

Figure 8. Map of the $PM_{2.5}$ concentration.

What is also worth noting is the fact that the level of air pollution in parks located in the centre of the city (Regent's Park or Hyde Park) is similar or even higher than less green areas further from the centre. Furthermore, as suggested by Milton and Steed (2007), the areas of high air pollution levels are often the same areas where the GPS signal is likely to be altered due to the urban-canyon effect, hence the location errors were found in the areas characterised by high air pollution.

Figure 9 shows the heatmap of all 195 activities completed by the research's participants. What can be easily noticed, the volunteers were active the most often in the Central London, particularly in the areas of Gower Streets and Regent's Park. This spatial pattern can be explained by the academic affiliation of the participants; the fact that almost all of the tracks have been completed by the members of the UCL Geography Running Club also justifies a very high frequency of the runs nearby Regent's Park and Primrose Hill.

6.1 Activities summary

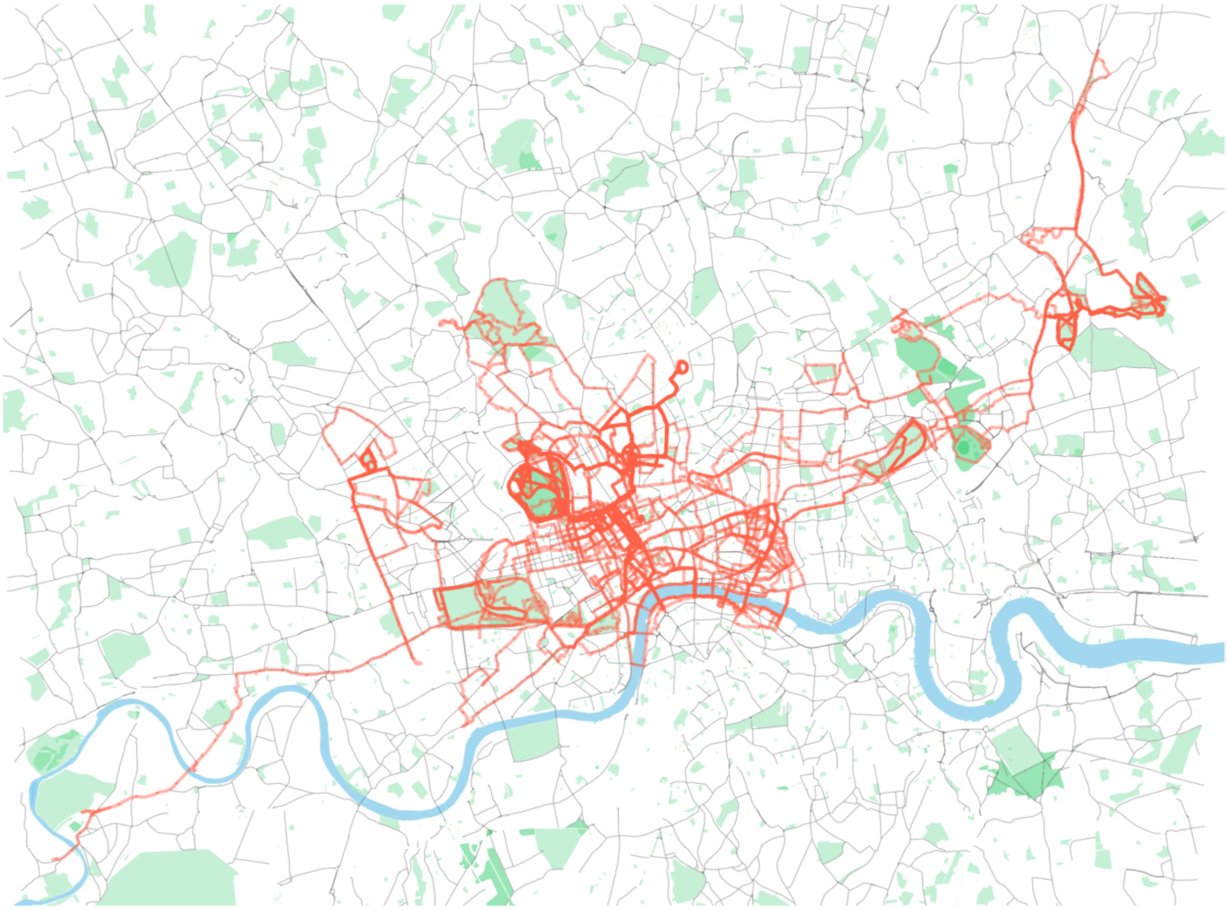

Figure 9. Heatmap of activities undertaken by the participants of the research.

The other thing that can be read from the map is that all participants are presumed to live on the Northern side of Thames (hence no activity on the Southern part of the city); moreover, two standing out heat spots in the areas of Stratford and Kensal Rise suggest possible neighbourhoods of residence. Although the information on the residence of the participants are not particularly relevant for this research, it clearly shows the potential of the activity data in the analysis of human mobility.

Table 2 presents descriptive statistics of the aggregation of all 195 completed activities.

Variable	Minimum value	1st Quantile	Median	Mean	3rd Quantile	Maximum value	Standard deviation
Speed (km/h)	0	11.27	14.57	14.88	18.21	30	5.59
Elevation (GPS Strava)	0.2	24.5	33	32.39	37.7	138	13.56
Slope (GPS Strava)	-148.07	-1.36	-0.01	-0.04	1.25	149.9	5.46
Elevation (Google API)	-1.285	19.96	27.07	28.17	34.52	134.77	14.15
Slope (Google API)	-142.87	-1.6	-0.09	-0.03	1.57	144.93	3.76
Lowess slope (Google API)	-103.22	-1.36	-0.01	-0.04	1.32	72.07	3.14
Lowess absolute slope (Google API)	0	0.94	1.75	2.45	3.14	103.22	2.4
Ventilation	1	1.047	1.088	1.091	1.12	2	0.06
Time difference	1	2	2	2.05	2	1993	12.4

Table 2. Descriptive statistics of all physical activities.

Both slopes derived from Strava and Google API data show that the runners most frequently moved on flat ground, with a median and mean extremely close to 0. Moreover, the results show that the time difference between subsequent points lasted 2 seconds in most of the cases, with outliers representing pauses in the activity; nevertheless, these outliers are crucial as they significantly contribute to the inhaled doses of air pollution. Notably, the results support favouring elevation data derived from Google API rather than relying on data collected by GPS devices; in spite of slightly higher standard deviation of Google API elevation, the standard deviation of API-origin slope is substantially lower than slope calculated from Strava data.

6.2 Runs and air pollution

Notably, elevation and slope were found to be significantly correlated with both the ventilation rate and $PM_{2.5}$ levels. In this sense, the pollutant is negatively correlated with elevation and smoothed absolute values of slope, suggesting that more elevated areas tend to have lower concentration of air pollution; this is optimistic considering the fact that such areas are characterized by increased ventilation and pollutant intake rates (see figure 10).

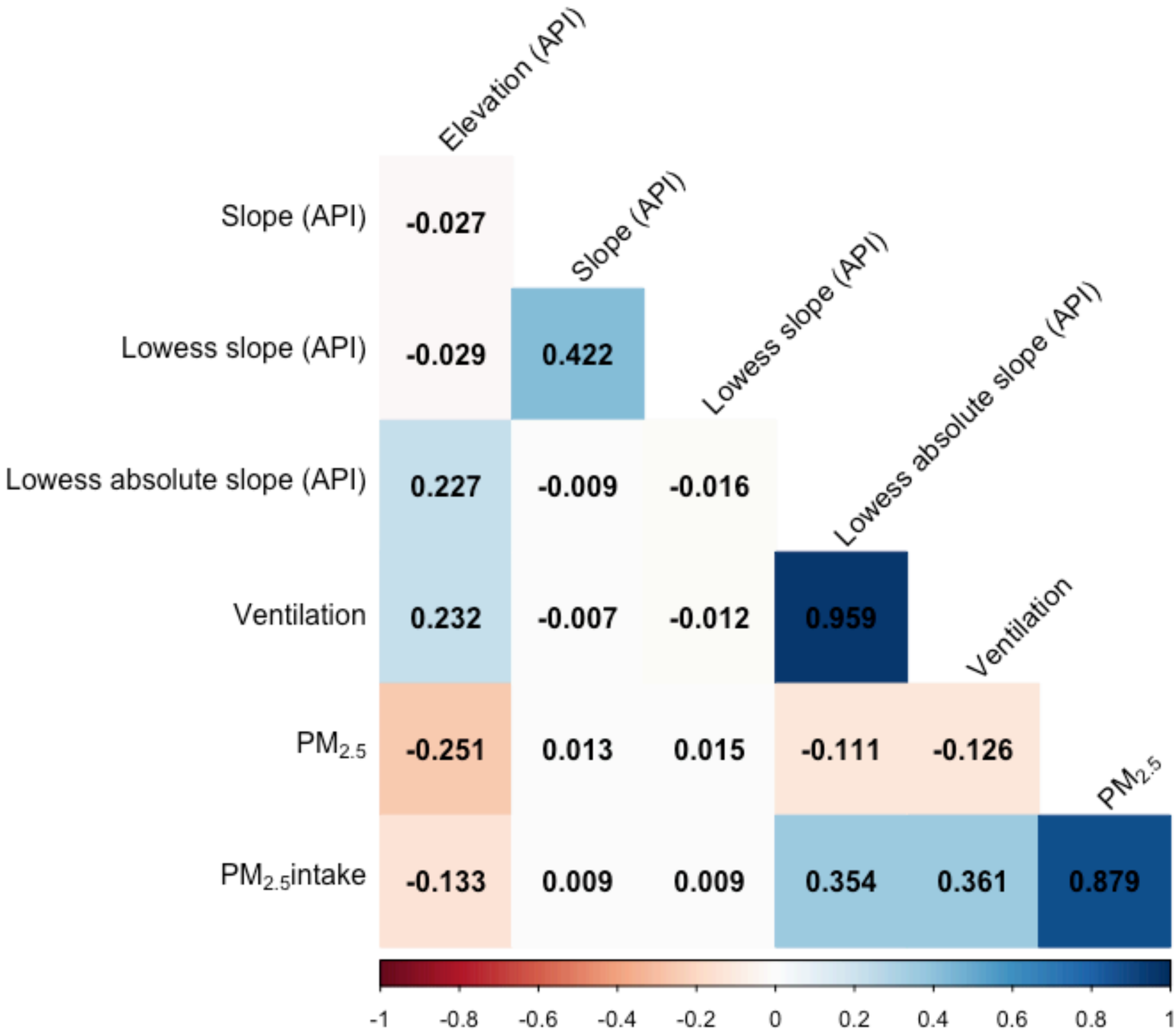

Figure 10. Correlation matrix of PM2.5 and elevation, slope and ventilation.

All of the correlation coefficients are highly statistically significant at p-value = 0.01. Unsurprisingly, the correlation between the concentration and intake of $PM_{2.5}$ is positive, statistically significant and extremely high. Nonetheless, the hypothesis that there is no difference in means of $PM_{2.5}$ concentration and $PM_{2.5}$ intake is rejected by the t-test, with p-value extremely close to 0; the means of concentration and intake of $PM_{2.5}$ equal 17.908 and 19.519, respectively.

Figure 11 presents the exposure rate of the participants calculated using the adopted model. By comparing these two maps, one should be able to spot places that are characterized by an increased exposure rate to air pollution, which would be a sign for a non-air-friendly route.

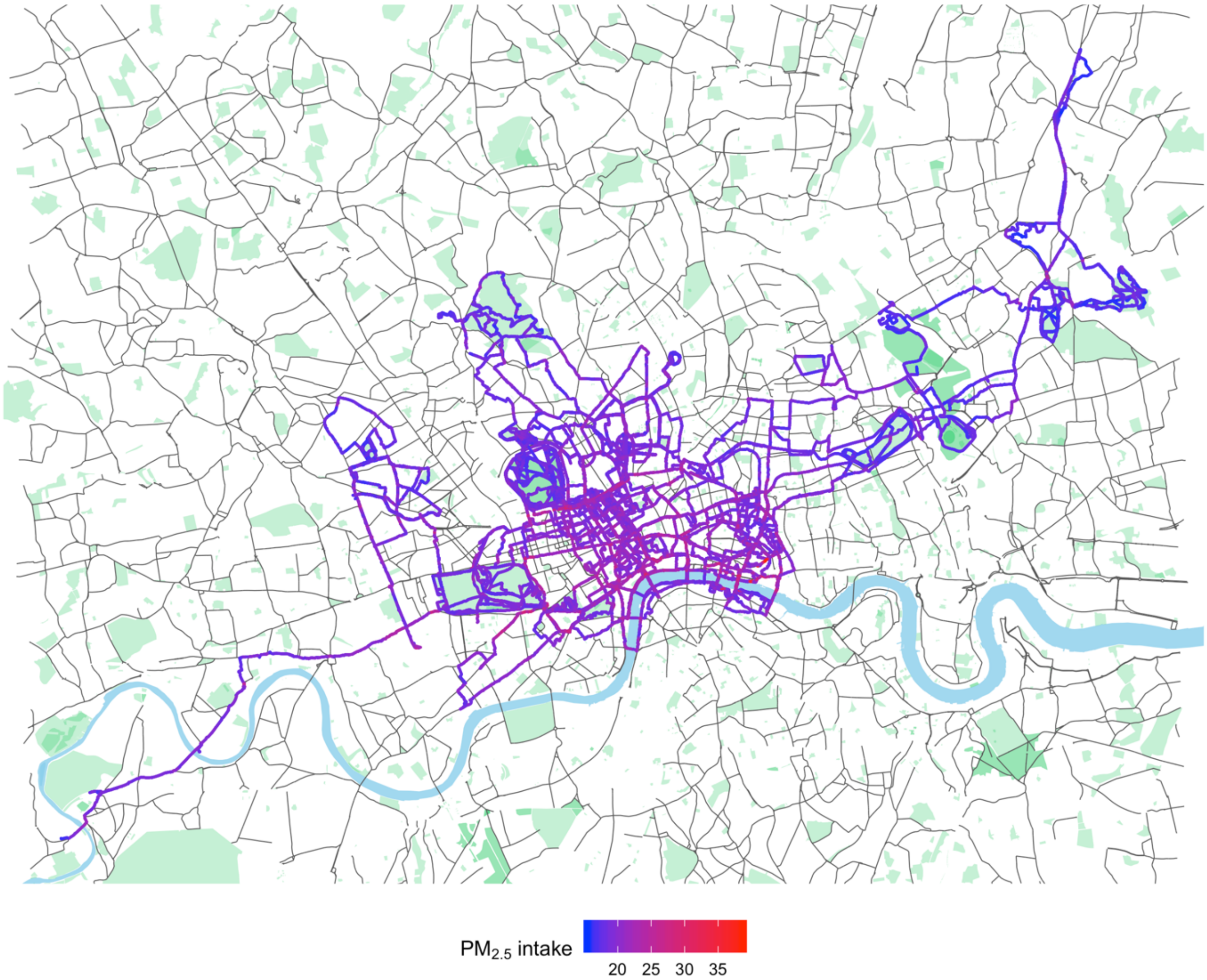

Figure 11. Map of the $PM_{2.5}$ intake.

6.3 Time-integrated exposure

However, looking on the aggregated records of pollutant intake might not be particularly useful considering the spatial scale of all activities. Therefore, the further analysis will focus on a few particularly interesting activities:

- Run A – most frequently chosen route that runs along a major road, Regent's Park, up to the Primrose Hill (moderate ascent);
- Run B – route that is characterized by the highest elevation and ascent (goes through Hampstead Heath);
- Run C – a synthetic run that has been produced by combining two similar runs from UCL to Regent's Park

6.3.1 Run A

Figure 12 presents the map for the $PM_{2.5}$ intake for Run A. While the highest amount of inhaled $PM_{2.5}$ is still nearby the major road, one can also spot an increased intake in the area of Primrose Hill. The $PM_{2.5}$ intake is still lower than the intake along the Euston Road, but is substantially higher than when running in Regent's Park.

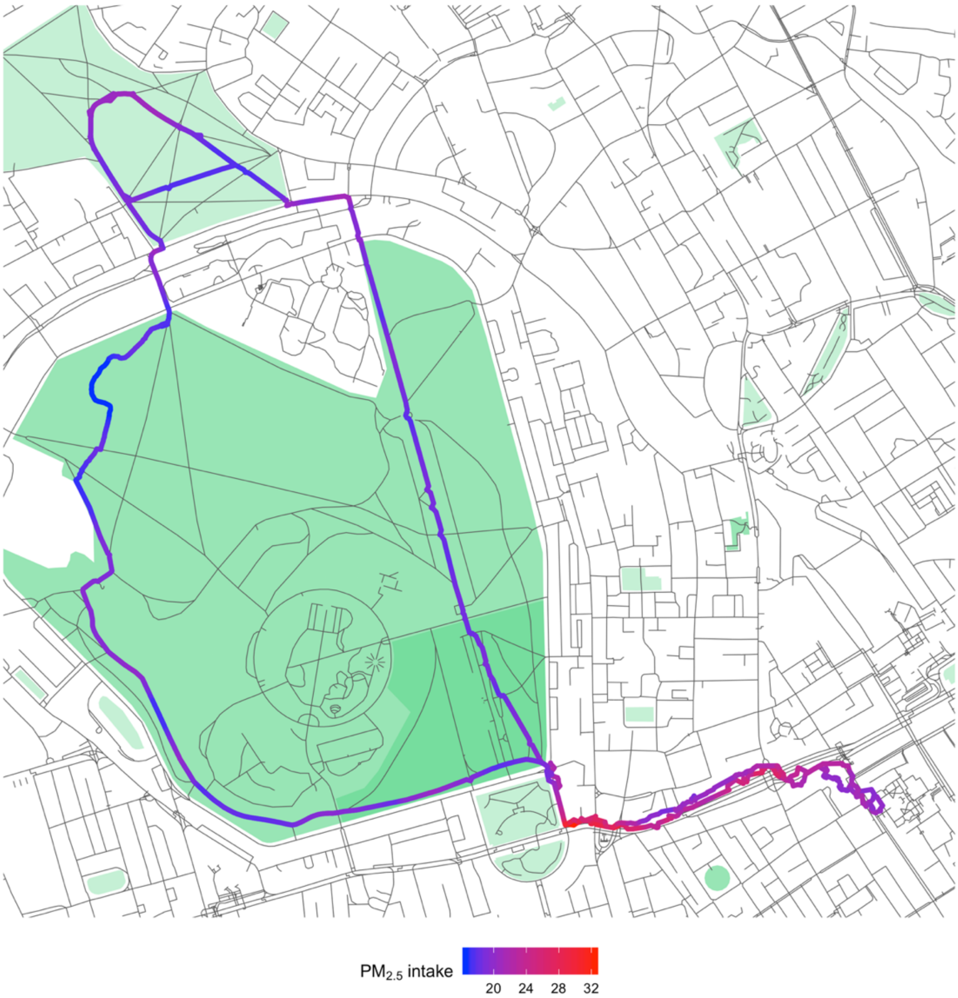

Figure 12. Map of the $PM_{2.5}$ intake for Run A.

The analysis of the first 5 minutes of Run A shows the difference between running on a road and in a park. Figures 13 and 14 presents the output of the time-integrated exposure model for two different segments of the same run.

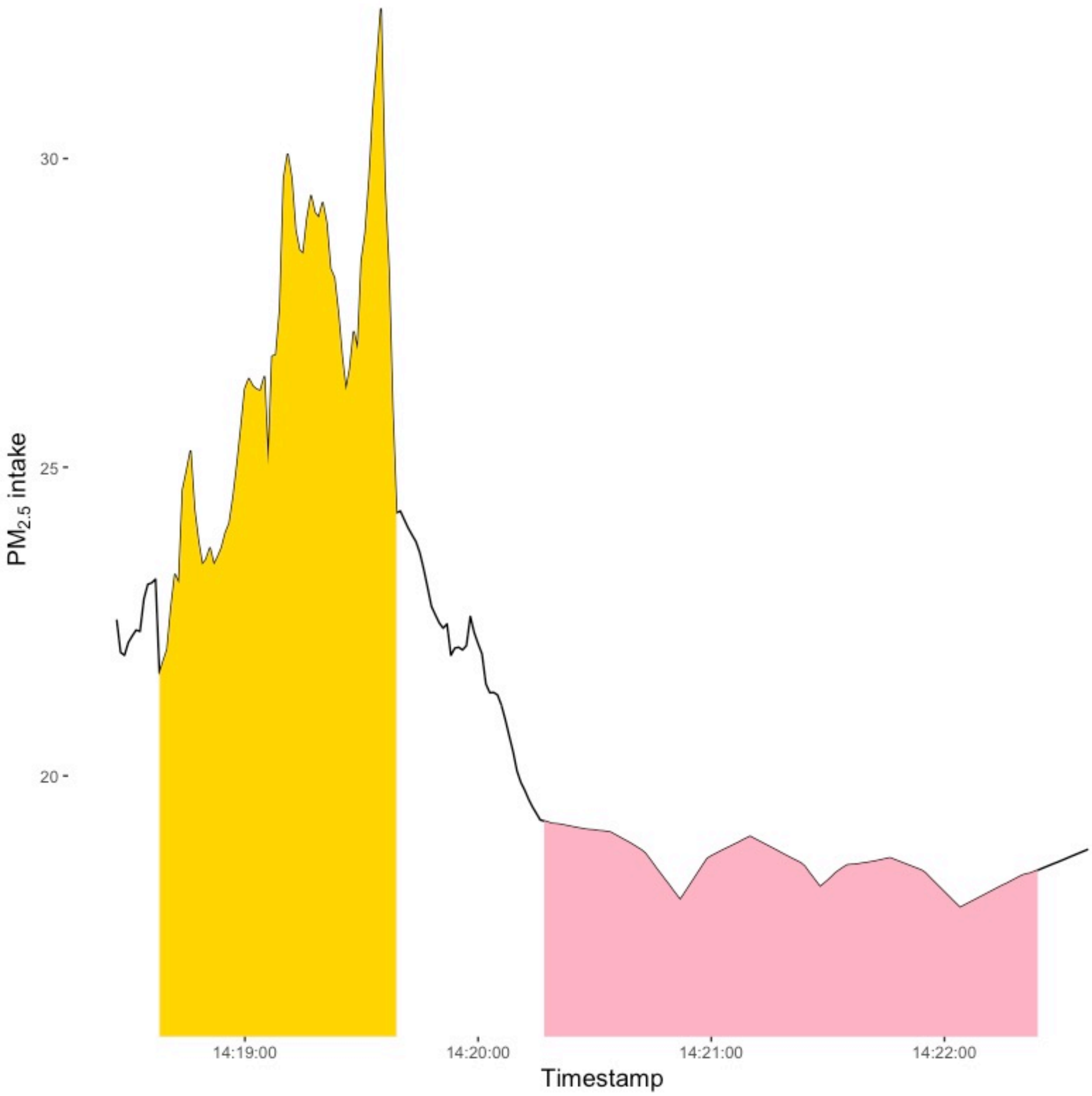

Figure 13. Graph showing time-integrated exposure for the beginning of Run A.

The calculations show that both segments are characterized by very similar time-integrated exposure (3863.05 and 3864.77), but different journey-time (143 seconds vs 208 seconds). In this sense, the result show that even though the road-based segment was over a minute shorter, a runner inhaled the same amount of pollutants as in a park; this is meaningful as it clearly shows the penalty of running to parks via congested roads and the importance of time in modelling exposure to air pollution.

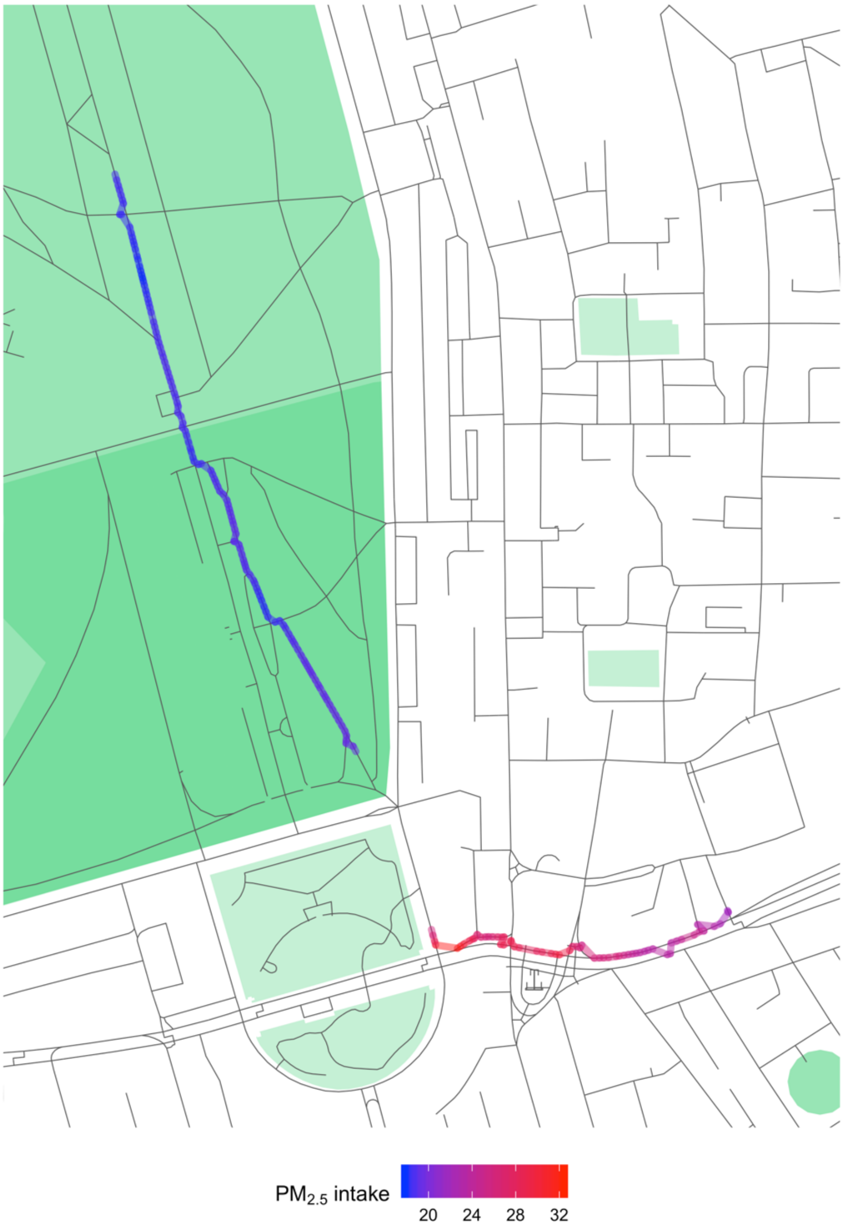

Figure 14. Map of two segments of Run A that are characterized by almost identical amounts of inhaled $PM_{2.5}$.

To take another route characterized by a similar course, figures 15 and 16 show how pollution intake when commuting from UCL to Regent's Park can be counterbalanced by running in the park. In this sense, while the time-integrated exposures of both fragments are comparable (9077 and 9072), the journey in the park was longer by over a minute and a half. During this time, the runner has run over 600 meters further. In other words, after running from UCL to Regent's Park, one can run for over 90 seconds until they reach the amount of pollution inhaled during the earlier part.

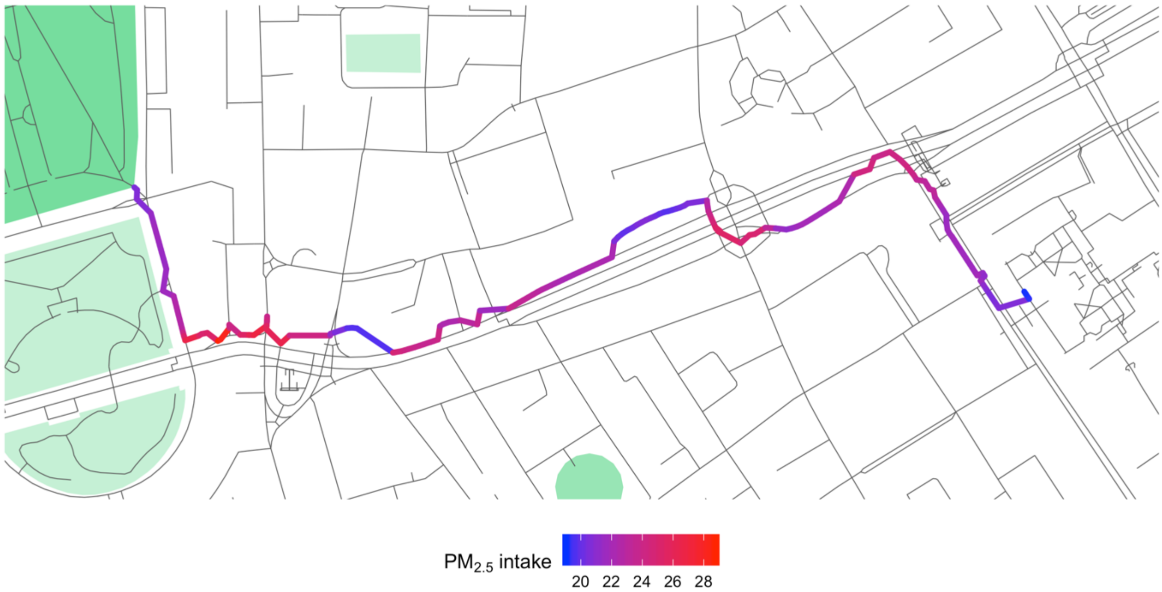

Figure 15. Map of the $PM_{2.5}$ intake for fragment X.

Figure 16. Map of the $PM_{2.5}$ intake for fragment Y.

6.3.2. Run B

The output of the time-integrated exposure for Run B underlines the importance of the duration of the activity in evaluating exposure to air pollution (see figure 17). The activity can be split into two segments (M and N), that both starts (bottom-right corner) and ends and the same place (Parliament Hill viewpoint).

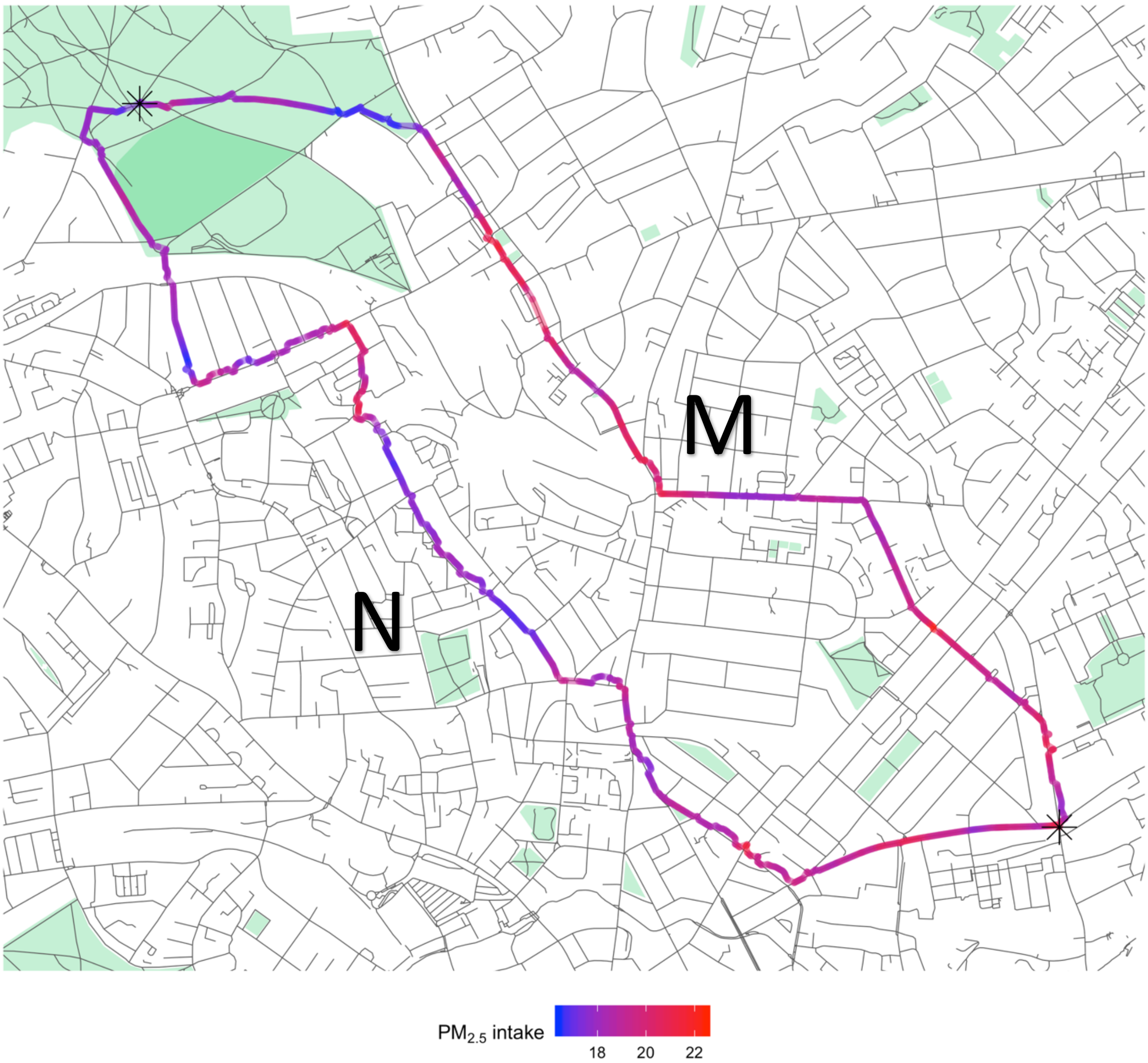

Figure 17. Map of the $PM_{2.5}$ intake for Run B.

Table 3 presents the output for both segments. While segment M was undertaken on busy roads, segment N is characterized by lower traffic volume, hence lower average pollution intake. Nevertheless, what is important is the fact that in spite of the lower mean pollution level, segment N is linked with a higher amount of inhaled $PM_{2.5}$ due to a longer duration of the activity.

	Time	Distance	Mean intake	Time-integrated exposure
Segment M	21:32 min	4746 m	19.13	24703
Segment N	24:18 min	5606 m	18.54	27039

Table 3. Output of Run B.

6.3.3 Run C

Finally, the results of time-integrated exposure for Run C show the significance of choosing main roads for a running route. Run C was derived by combining two activities that have shared a majority of undertaken route. This hypothetical run is estimated to last for 30 minutes, covering almost 7.8 kilometres (see figure 18).

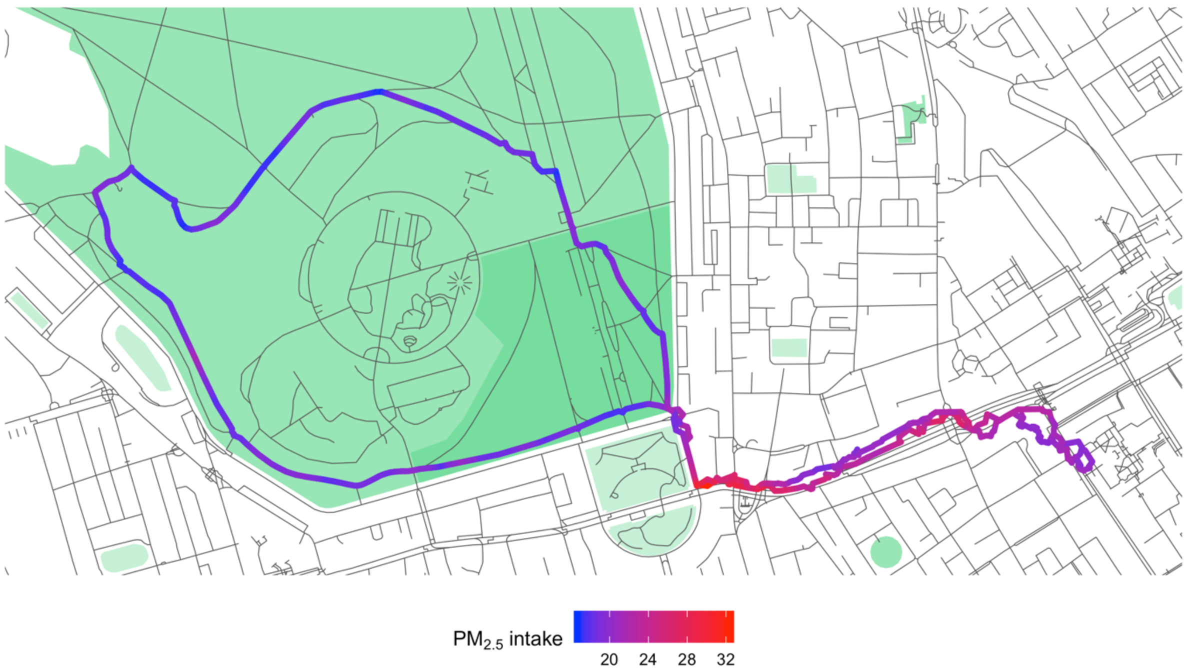

Figure 18. Map of the $PM_{2.5}$ intake for Run C.

During that run, the time-integrated exposure of a runner would equal 38765.63. Whereas it is not very meaningful, what is interesting is the proportion of the exposure along the route. In this sense, in spite running to and back from Regent's Park via Euston Road (figure 19) would be almost 2 minutes and over 700 meters shorter than the activity in the park (figure 20), it would be associated with much higher amount of inhaled $PM_{2.5}$.

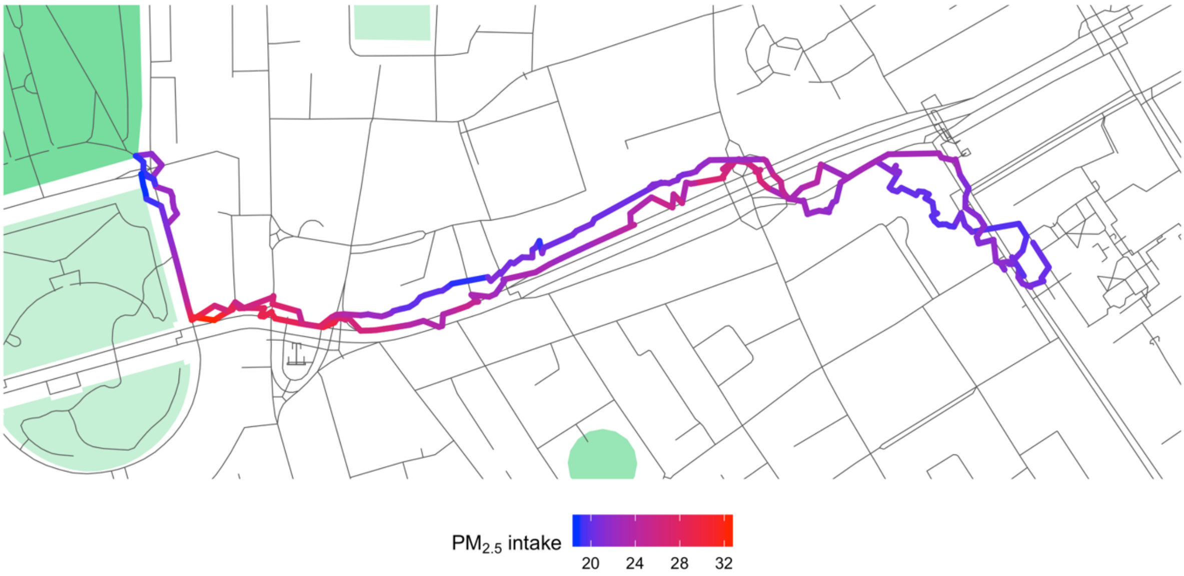

Figure 19. Map of the PM$_{2.5}$ intake for the beginning and finish of Run C.

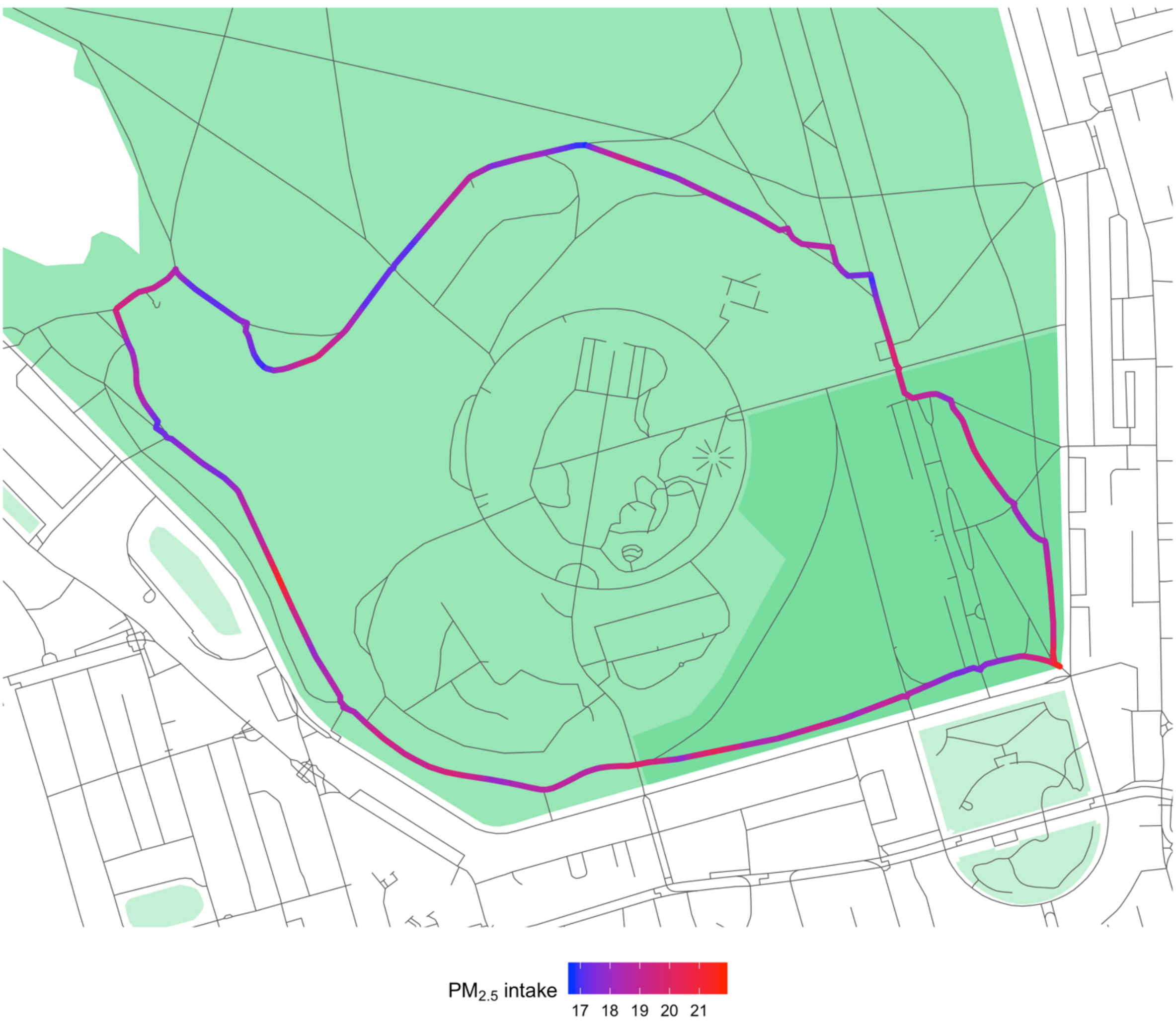

Figure 20. Map of the PM$_{2.5}$ intake the middle part of Run C.

7 Discussion

The presented results suggest that it does actually matter how runners choose their running routes. It is apparent that the concentration of air pollution along the busy roads is extremely high and harmful. Correspondingly, the results suggest that the airborne pollution intake while running along the major road is much higher than in parks, in spite of the difference in the ventilation rates linked with increased inclination in parks. In other words, despite increased ventilation due to more hilly roads, the inhalation rate in parks is still substantially lower than when running along busy roads.

Moreover, it was found that the duration of journey has a much greater effect on the exposure to air pollution, than the ventilation rate – the longer we run, the more airborne pollutants we inhale. This means that in order to mitigate the health risks caused by being exposed to high levels of air pollution, one should try to shorten the duration of an activity. All of the examples suggest that longer tracks characterized by lower air pollution are associated with a higher amount of inhaled pollutants than more polluted but shorter and faster tracks.

These findings are crucial in terms of how runners can choose their running routes. In this sense, it is argued that, runners who do not care about the elapsed distance or duration of an activity should choose less busy and thus less polluted roads; in fact, running along back streets is associated with a decrease of exposure to pollution that is caused by a heavy traffic (King’s College London, 2018c). This also means that such runners are advised to spend most of their activity running in a park, regardless the elevation profile of the route. Respectively, runners who run for the purpose of getting from point A to point B and for whom time matters (for instance people who commute by running and therefore understand it as a mean of transport) should choose routes as fast and short as possible. Such routes allow them to inhale the least amount of airborne pollutants by reducing the duration of being exposed to air pollution.

While it is understood that it might be difficult to influence the path decisions of runners, as runners might simply not think about it during the run, it is also believed that such studies may be of policy makers’ interest, who by their decisions can influence the society. In this sense, by facilitating citizens with policies that could make them exercise more often, it is assumed that the greater knowledge on air-friendly routes and exposure rates can result in better understanding of the topic and thus raise awareness of air pollution, its consequences and potential preventive measures that could be taken.

By developing and spreading the knowledge about air-friendly routes, policy makers can empower people to evaluate their daily exposure to air pollution and mitigate their exposure to air pollution by enabling them to change their daily routines (Huck *et al,* 2017). Furthermore, considering the fact that "individuals constantly exploit a small set of repeatedly visited location", it is believed that people rarely change their mobility patterns (Alessandretti *et al,* 2018: 485).

This means that by educating urban citizens on their exposure to air pollution, policy makers can promote physical activity in urban dwellings. In fact, as most of journeys are completed on foot by citizens, promoting running as a utilitarian mode of transport is understood to result in the improvement of health of citizens and reduction of obesity across population (Winters *et al,* 2007).

What is more, promotion of running as a method of daily commute is also believed to be one of the most effective ways of fighting with high levels of air pollution in cities. By running, walking or cycling to work, instead of using a public transport, citizens can minimise their ecological footprint thus significantly reducing their contribution to air pollution. It is hoped that implementing policies restricting car travels in cities' centres would foster running in cities.

7.1 Potential applications

As the study investigates the time-integrated exposure to average level of air pollution of runners whose activities have been recorded with a GPS device, the research is easily applicable to other cities and users of other fitness applications, such as Nike+ Run Club. As London is a relatively flat city, the examination of more hilly cities could reveal some new interesting patterns.

Moreover, knowing that the choice of a route is related to the time-integrated exposure to air pollution, it is believed that enabling people to assess their exposure gives them opportunity to predominantly plan their running routes based on the most air-friendly options. In this sense, this dissertation is also hoped to serve as a pilot study for projects that could use pathfinding algorithms to predict and find the best possible routes; such method helps to find the most optimal way based on variables of interests, in this case air pollution. In spite of the great potential, the framework for the application of the technique in everyday life is still under development. Although there are some start-ups, such as RunFriendly that focus on this particular topic of air-friendly routes, more public and private organisations could potentially focus on addressing the problem of air pollution using the approach undertaken in this dissertation.

7.2 Further research

This research has investigated a few runners who use Strava to record their activities and their exposure to air pollution in London. It is recommended that future research will use the proposed framework to evaluate exposure to air pollution for the entire population. Knowing that the effect of exposure to air pollution might vary between individuals, more data are needed to enhance the credibility of the findings (An *et al,* 2018). In this sense, it is hoped that future studies that would have access to much more robust and comprehensive datasets on human mobility, potentially from a wide range of sources, could increase the internal and external validity of the research by analysing a wider proportion of the population over the whole city.

Moreover, as the ventilation rates and time-integrated exposure to air pollution is understood to be heterogenous for various types of physical activity, including cycling and walking, it is recommended for the future research to include in the analysis all other forms of outdoor physical activity (Lu and Fang, 2015). In fact, it is believed that the type of activity does matter as it significantly affects one's ventilation rate and duration of the exercise; for instance, cycling tends to be characterized by shorter duration of an activity and lower ventilation rate, but at the same time cyclists usually move along more busy and polluted roads. While such widening of the scope of the research is desired, it also needs to be recognized that those activities differ in terms of the associated exposure to air pollution, meaning that each of them should be studied individually.

Finally, future research could evaluate short-term and long-term changes in the level of air pollution. Considering the fact that concentration and exposure to air pollution may differ depending on the current weather conditions, including temperature and humidity, it is advised to study how these changes might affect the choice of routes (Smith *et al,* 2017). In this sense, alternative methods such as machine learning can be used in the analysis to improve the assessment of health implications of the exposure to air pollution (Bellinger *et al,* 2017). Furthermore, considering increasing concern about climate change, but also general awareness of the necessity of reducing air pollution in urban dwellings, future research could evaluate how the exposure to air pollution has changed over the years and to what extent the current predictions of annual averages of air pollution will come true. Importantly, as it is planned to introduce the Ultra Low Emission Zone (ULEZ) in London, the level of air pollution in the central London is predicted to decrease by almost 50% (Greater London Authority, 2019). This means that the spatial distribution of the pollution concentration may become more even, resulting in main roads being less harmful for runners; this is important as it may change the way runners should choose their running routes.

Auto-Critique

The main objective of this dissertation was to evaluate exposure to air pollution of runners in London; this objective has been met partially. The dissertation initially was planned to include data facilitated by Strava, however, as noted in Chapter 4, the study has eventually focused on data from a few volunteers, meaning that the exposure to air pollution was calculated only for a small subset of the population.

The issues with the quantity of data means that the dissertation does not enable to generalize the findings to the wider population, meaning that the external validity of the research is low. Nonetheless, it is believed that more data on physical activities in urban dwelling should solve this concern.

The greatest weakness of the study lies in the assumptions linked to the calculation of the ventilation rate. As identified in Chapter 3.3.1, the ventilation rate model assumes that ventilation is independent from time and individual characteristics. However, whereas the assumption of the homogeneity of population threatens the external validity of the research as the respiratory patterns of individuals differ, the assumption about no-time effect on the ventilation rate is a serious problem for the internal validity of the study. These two assumptions are essential for the application of this simple model in this work, nevertheless they greatly affect the precision and reliability of the research. In spite of the fact that the dissertation helps to understand how time and inclination influence exposure to air pollution, further research is needed to more accurately model changes in inhalation and ventilation considering the effects of time and personal background. Moreover, to improve the ventilation rate model, data on the respiratory patterns of runners could be used; this could result in the improvement of the consistency and accuracy of the results.

Additionally, as the dissertation has focused on London, other cities should be investigated considering variations in air pollution, traffic volume and elevation profile. In this sense, while London is characterized by high air pollution, high traffic volume and moderate variations in elevation profile, there are some European cities that are flatter and have much higher levels of air pollution and traffic volume, meaning that the time-integrated exposure to air pollution might be significantly different. In this sense, the results of this study should not be applied to other cities.

Despite the limitations and weaknesses of the research that are the result of limited access to data, this dissertation is innovative as it provides a new approach to the investigation of the exposure to air pollution of urban citizens. As it examines long-term time-integrated exposure to air pollution and real-time movement of runners, it suggests a framework that allows to make predictions about the exposure rate and suggest how runners should run in London in order to minimise their exposure to air pollution.

Appendix A

This appendix includes information on data that have been used in this research. Elevation data were obtained using Google Elevation API.

File	Description	Source
PostLAEI2013_2013_NO2.csv	Modelled annual average concentration of NO_2 at 20m grid level for 2013.	Greater London Authority https://data.london.gov.uk/dataset/london-atmospheric-emissions-inventory-2013
PostLAEI2013_2013_NOx.csv	Modelled annual average concentration of NO_x at 20m grid level for 2013.	
PostLAEI2013_2013_PM10.csv	Modelled annual average concentration of PM_{10} at 20m grid level for 2013.	
PostLAEI2013_2013_PM10d.csv	Modelled annual average concentration of PM_{10d} at 20m grid level for 2013.	
PostLAEI2013_2013_PM25.csv	Modelled annual average concentration of $PM_{2.5}$ at 20m grid level for 2013.	
modelled_pollutants.csv	The file contains merged modelled annual average concentrations of the pollutants.	
run.pollution.csv	The file contains information of all runs including predicted values of modelled annual average concentrations of the pollutants for each data point.	

Folder	Description	Source
Strava	The folder contains 221 records of Strava activities that have been collected. Some of the files were removed due to the fact that they were not completed in London.	All of the files were collected from the volunteers of this research.
Shapefiles	The folder contains a shapefile for London boundaries at the Ward level.	Greater London Authority https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london
MyShapefiles	The folder contains all shapefiles that were created in the process of the analysis to simplify and short-circuit the calculations.	
Scripts	The folder contains all R scripts for the analysis conducted within this dissertation.	

List of References

5kevents.org (n.d.) 'Runner demographics – Statistics and Research' (WWW) Racine, WI: 5kevents.org. Available at: https://www.5kevents.org/demographics.html (Accessed 15 March 2018).

Alessandretti, L., P. Sapiezynski, V. Sekara, S. Lehmann and A. Baronchelli (2018) 'Evidence for a Conserved Quentity in Human Mobility', *Nature Human Behaviour,* 2, pp. 485-491.

Ali, O., F. Di Nardo, A. Harrison and A. Verma (2017) 'The link between perceived characteristics of neighbourhood green spaces and adults' physical activity in UK cities: analysis of the EURO-URHIS 2 Study', *The European Journal of Public Health,* 27, 4, pp. 761-765.

An, R., S. Zhang, M. Ji and C. Guan (2018) 'Impact of ambient air pollution on physical among adults: a systematic review and meta-analysis', *Perspectives in Public Health,* 138, 2, pp. 111-121.

Barstow, T. J. and P. A. Mole (1991) 'Linear and nonlinear characteristics of oxygen uptake kinetics during heavy exercise', *Journal of Applied Physiology,* 71, pp. 2099-2106.

Baty, F., C. Ritz, A. van Gestel, M. Brutsche and D. Gerhard (2016) 'Modeling the oxygen uptake kinetics during exercise testing of patients with chronic obstructive pulmonary diseases using nonlinear mixed models', *BMC Medical Research Methodology,* 16, 66, pp. 1-9.

Bellinger, C., M. S. M. Jabbar, O. Zaiane and A. Osornio-Vargas (2017) 'A systematic review of data mining and machine learning for air pollution epidemiology', *BMC Public Health,* 17, pp. 907-926.

Bergstrom, H. C., T. J. Housh, K. C. Cochrane, N. D. M. Jenkins, S. L. Buckner, J. A. Goldsmith, J. M. Zuniga, R. J. Schmidt, G. O. Johnson and J. T. Cramer (2015) 'Application of the Critical Heart Model to Treadmill Running', *Journal of Strength and Conditioning Research,* 29, 8, pp. 2237-2248.

Berry, M. J., C. J. Dunn, C. L. Pittman, W. C. Kerr and N. E. Adair (1996) 'Increased ventilation in runners during running as compared to walking at similar metabolic rates', *European Journal of Applied Physiology,* 73, pp. 245-250.

Bonsingore, M. R., G. Morici, P. Abate, S. Romano and G. Bonsignore (1998) 'Ventilation and entrainment of breathing during cycling and running in triathletes', *Medicine & Science in Sports & Exercise,* 30, 2, pp. 239-245.

Brauer, M. (2010) 'How Much, How Long, What, and Where: Air Pollution Exposure Assessment for Epidemiologic Studies of Respiratory Disease', *Proceedings of the American Thoracic Society,* 7, pp. 111-115.

Buechler, J. (2018) 'Your exposure to air pollution could be much higher than your neighbour's – here's why', *The Conversation,* 22 June. Available at: https://theconversation.com/your-exposure-to-air-pollution-could-be-much-higher-than-your-neighbours-heres-why-98486 (Accessed 20 February 2019).

Carter, K. (2015) 'Marathons by numbers: running the data', *The Guardian,* 21 April. Available at: https://www.theguardian.com/lifeandstyle/the-running-blog/2015/apr/21/marathons-by-numbers-running-the-data (Accessed 15 March 2018).

Cepeda, M., J. Schoufour, R. Freak-Poli, C. M. Koolhaas, K. Dhana, W. M. Bramer and O. H. Franco (2017) 'Levels of ambient air pollution according to mode of transport: a systematic review', *Lancet Public Health,* 2, pp. 23-34.

Cleveland, W. S. (1981) 'LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression', *The American Statistician,* 35, 1, pp. 54.

Cook, S., J. Shaw and P. Simpson (2016) 'Jography: Exploring Meanings, Experiences and Spatialities of Recreational Road-running', *Mobilities,* 11, 5, pp. 744-769.

De Brabandere, A., T. Op De Beeck, K. H. Schutte, W. Meert, B. Vanwanseele and J. Davis (2018) 'Data fusion of body-worn accelerometers and heart rate to predict VO_{2max} during submaximal running', *PLoS ONE,* 13, 6, pp. 1-17.

De Nazelle, A., E. Seto, D. Donaire-Gonzalez, M. Mendez, J. Matamala, M. J. Nieuwenhuijsen and M. Jerret (2013) 'Improving estimates of air pollution exposure through ubiquitous sensing technologies', *Environmental Pollution,* 176, pp. 92-99.

De Smet, D., M. Francuax, J. M. Hendrickx and M. Verleysen (2016) 'Heart rate modelling as a potential physical fitness assessment for runners and cyclists', *Proceedings of the Machine Learning and Data Mining for Sports Analyitcs Workshop at ECML/PPKD,* Riva del Garda, Italy, September 2016, pp. 1-8.

Dias, D. and O. Tchepel (2014) 'Modelling of human exposure to air pollution in the urban environment: a GPS-based approach', *Environmental Science and Pollution Research,* 21, pp. 3558-3751.

Dias, D. and O. Tchepel (2018) 'Spatial and Temporal Dynamics in Air Pollution Exposure Assessment', *International Journal of Environmental Research and Public Health,* 15, 558, pp. 1-23.

Eaton, M. D., D. L. Evans, D. R. Hodgson and R. J. Rose (1995) 'Effect of treadmill inline and speed on metabolic rate during exercise in Thoroughbred horses, *Journal of Applied Physiology,* 79, 3, pp. 951-957.

Evans, M. R., D. Oliver, X. Zhou and S. Shekhar (2014) 'Spatial Big Data: Case Studies on Volume, Velocity, and Variety, in Karimi, H. A. [ed.] *Big Data: Techniques and Technologies in Geoinformatics,* London: CRC Press, pp, 149-176.

Gerharz, L. E., O. Klemm, A. V. Broich and E. Pebesma (2013) 'Spatio-temporal modelling of individual exposure to air pollution and its uncertainty', *Atmospheric Environment,* 64, pp. 56-65.

Glaeser, E. (2011) 'Our Urban Species', in Glaeser, E. [ed.] *The Triumph of the City,* New York: Penguin, pp. 1-15.

Greater London Authority (2019) 'The Mayor's Ultra Low Emission Zone for London' (WWW) London: Greater London Authority. Available at: https://www.london.gov.uk/what-we-do/environment/pollution-and-air-quality/mayors-ultra-low-emission-zone-london (Accessed 11 March 2019).

Guinness World Record (2019) ‘Fastest run 100 meters (male)’ (WWW) Available at: http://www.guinnessworldrecords.com/world-records/fastest-run-100-metres-(male) (Accessed 9 March 2019).

Gulliver, J. and D. J. Briggs (2005) ‘Time-space modelling of journey-time exposure to traffic-related air pollution using GIS’, *Environmental Research,* 97, pp. 10-25.

Haddad, A. Y. Zhang, S. Tu, B. Celler and H. Nguyen (2014) ‘Modelling and regulating of cardio-respiratory response for the enhancement of interval training’, *BioMedical Engineering OnLine,* 13, 9, pp. 1-14.

Hanson, N. and J. Buckworth (2017) ‘Personality characteristics of barefoot runners: openness and conscientiousness as the defining traits’, *Sports Science Health,* 13, pp. 33-38.

Hern, A. (2018) ‘Strava suggests military users ‘opt out’ of heatmap as row deepens’, *The Guardian,* 29 January. Available at: https://www.theguardian.com/technology/2018/jan/29/strava-secret-army-base-locations-heatmap-public-users-military-ban (Accessed 14 March 2018).

Hitchings, R. and A. Latham (2016) ‘Indoor versus outdoor running: understanding how recreational exercise comes to inhabit environments through practitioner talk’, *Transactions of the Institute of British Geographers,* 41, 4, pp. 503-514.

Hitchings, R. and A. Latham (2017) ‘Exercise and environment: New qualitative work to link popular practice and public health’, *Health and Place,* 46, pp. 300-306.

Hoffman, M. D. and K. Fogard (2012) ‘Demographic Characteristics of 161-km Ultramarathon Runners’, *Research in Sports Medicine,* 20, 1, pp. 59-69.

Huck, J. J., J.D. Whyatt, P. Coulton, B. Davison and A. Gradinar (2017) ‘Combining physiological, environmental and locational sensors for citizen-oriented health applications’, *Environmental Monitoring and Assessment,* 189, pp. 114-128.

Hunt, K. J. and A. J. R. Hunt (2016) 'Feedback control of heart rate during outdoor running: A smartphone implementation', *Biomedical Signal Processing and Control,* 26, pp. 90-97.

Kamargianni, M. (2015) 'Investigating next generation's cycling ridership to promote sustainable mobility in different types of cities, *Research in Transportation Economics,* 53, pp. 45-55.

Karsky, D. (2004) 'Comparing Four Methods of Correcting GPS Data: DGPS, WAAS, L-Band, and Postprocessing', *Technology and Development Program,* 7100, pp. 1-6.

King's College London (2018a) 'About londonair' (WWW) London: King's College London. Available at: https://www.londonair.org.uk/LondonAir/General/about.aspx (Accessed 12 March 2019).

King's College London (2018b) 'Annual Pollution Maps' (WWW) London: King's College London. Available at: https://www.londonair.org.uk/london/asp/annualmaps.Asp?Species=no2 (Accessed 5 September 2018).

King's College London (2018c) 'What can I do to avoid air pollution' (WWW) London: King's College London. Available at: https://www.londonair.org.uk/LondonAir/guide/MyActionsForMe.aspx (Accessed 20 February 2019).

Kim, S., E. P. Homestead and W. C. Byrnes (2018) 'Breathing valve resistance alters physiological responses during a graded exercise test', *European Journal of Applied Physiology,* 118, pp. 1921-1929.

Latham, A. (2015) 'The history of a habit: jogging as a palliative to sedentariness in 1960s America', *Cultural Geographies,* 2015, 22, 1, pp. 103-126.

Latham, A. (2017) 'Mobility', in Latham, A., D. McCormack, K. McNamara and D. McNeill [eds] *Key Concepts in Urban Geography,* London: Sage, pp. 27-36.

Latham, A. and P. R. H. Wood (2015) 'Inhabiting infrastructure: exploring the interactional spaces of urban cycling', *Environment and Planning A,* 47, pp. 300-319.

Lewis, A. and P. Edwards (2016) ‘Validate personal air-pollution sensors’, *Nature,* 535, pp. 29-31.

Lindsay, G., A. Macmillan and A. Woodward (2011) ‘Moving urban trips from cars to bicycles: impact on health and emissions’, *Australian and New Zealand Journal of Public Health,* 35, 1, pp. 54-60.

Lipar, P., I. Strnad, M. Cesnik and T. Maher (2016) ‘Development of Urban Driving Cycle with GPS Data Post Processing’, *Promet – Traffic & Transportation,* 28, 4, pp. 353-364.

Liu, J., H. Zheng, T. Feng, S. Yuan and H. Lu (2013) ‘Post-processing procedures for passive GPS based travel survey’, *Procedia – Social and Behavioural Sciences,* 96, pp. 310-319.

Lu, Y. and Fang T. B. (2015) ‘Examining Personal Air Pollution Exposure, Intake, and Health Danger Zone Using Time Geography and 3D Geovisualization’, *ISPRS International Journal of Geo-Information,* 4, pp. 32-46.

Ludwig, M., M. Fuller, A. M. Sundaram and A. Asteroth (2015) ‘On Modeling the Cardiovascular System and Predicting the Human Heart Rate under Strain’, *In Proceedings of the 1st International Conference on Information and Communication Technologies for Aging Well and e-Health,* pp. 106-117.

Mapping London (n.d.) ‘Strava Labs Heatmap of Runners’ (WWW) Available at: http://mappinglondon.co.uk/2018/strava-labs-heatmap-of-runners/ (Accessed 11 March 2019).

Mead, M. I., O. A. M. Popoola, G. B. Stewart, P. Landshoff, M. Calleja, M. Hayes, J. J. Baldovi, M. W. McLeod, T. F. Hodgson, J. Dicks, A. Lewis, J. Cohen, R. Baron, J. R. Saffell and R. L. Jones (2013) ‘The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks’, *Atmospheric Environment,* 70, pp. 186-203.

Mila, C., M. Salmon, M. Sanchez, A. Ambros, S. Bhogadi, V. Sreekanth, M. Nieuwenhuijsen, S. Kinra, J. D. Marshall and C. Tonne (2018) ‘When, Where, and What? Characterizing Personal PM2.5 Exposure in Periurban India by Integrating GPS, Wearable Camera, and Ambient and Personal Monitoring Data, *Environmental Science and Technology,* 52, pp. 13481-13490.

Millet, G. P., V. E. Vleck and D. J. Bentley (2009) ‘Physiological Differences Between Cycling and Running: Lessons from Triathletes’, *Sports Medicine,* 39, 3, pp. 179-206.

Milton, R. and A. Steed (2005) ‘Correcting GPS Readings from a Tracked Mobile Sensor’, *Proceedings of the First International conference on Location- and Context-Awareness,* pp. 83-94.

Milton, R. and A. Steed (2007) ‘Mapping carbon monoxide using GPS tracked sensors’, *Environmental Monitoring and Assessment,* 123*,* pp. 1-19.

Minetti, A. E., L. P. Ardigo and F. Saibene (1994) ‘Mechanical determinants of the minimum energy of gradient running in humans’, *Journal of Experimental Biology,* 195, pp. 211-225.

Munn, E. (2018) ‘Best fitness apps: Best apps for running, cycling, calorie countring and more’, 28 February. Available at: https://www.expertreviews.co.uk/fitness-equipment/1406812/best-fitness-apps-running-cycling (Accessed 10 March 2019).

Nwokoro, C., C. Ewin, C. Harrison, M. Ibrahim, I. Dundas, I. Dickson, N. Mushtaq and J. Grigg (2012) ‘Cycling to work in London and inhaled dose of black carbon’, *European Respiratory Journal,* 40, 5, pp. 1091-1097.

Padulo, J., D. Powell, R. Milia and L. P. Ardigo (2012) ‘Kinematics of Running at Different Slopes and Speeds’, *Journal of Strength and Conditioning Research,* 26, 5, pp. 1331-1339.

Padulo, J., D. Powell, R. Milia and L. P. Ardigo (2013) ‘A Paradigm of Uphill Running’, *PLoS One,* 8, 7, pp. 1-8.

Piatkowski, D. P. and W. E. Marshall (2015) ‘Not all prospective bicyclists are created equal: The role of attitudes, socio-demographics, and the built environment in bicycle commuting’, *Travel Behaviour and Society*, 2, pp. 166-173.

Pucher, J., R. Buehler, D. R. Bassett and A. L. Dannenberg (2010) ‘Walking and Cycling to Health: A Comparative Analysis of City, State, and International Data’, *American Journal of Public Health,* 100, 10, pp. 1986-1992.

Robb, D. (2017) ‘Building the Global Heatmap’ (WWW), 1 November, *Medium.* Available at: https://medium.com/strava-engineering/the-global-heatmap-now-6x-hotter-23fc01d301de (Accessed 10 August 2018).

Running USA (2017) ‘2017 National Runner Survey’ (WWW) Los Angeles: Running USA. Available at: http://cdn.trustedpartner.com/docs/library/RunningUSA2012/RunningUSA_NRS_2017.pdf (Accessed 15 March 2018).

Sallis, J. F., E. Cerin, T. L. Conway, M. A. Adams, L. D. Frank, M. Pratt, D. Salvo, J. Schipperijn, G. Smith, K. L. Cain, R. Davey, J. Kerr, P. Lai, J. Mitas, R. Reis, O. L. Samiento, G. Schofield, J. Troelsen, D. Van Dyck, I. De Bourdeaudhuij and N. Owen (2016) ‘Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study’, *Lancet,* 387, pp. 2207-2217.

Schroter, R. C. and D. J. Marlin (2002) ‘Modelling the oxygen cost of transport in competitions over ground of variable slope’, *Equine Exercise Physiology,* 34, pp. 397-401.

Scott, R. A., E. Georgiades, R. H. Wilson, W. H. Goodwin, B. Wolde and Y. P. Pitsiladis (2003) ‘Demographic Charecteristics of elite Ethipian Endurance Runners’, *Medicine and Science in Sports And Exercise,* 35, 10, pp. 1727-1732.

Short, C. E., C. Vandelanotte, A. Rebar and M. J. Duncan (2014) ‘A Comparison of Correlates Associated With Adult Physical Activity Behavior in Major Cities and Regional Settings’, *Health Psychology,* 33, 11, pp. 1319-1327.

Smith, K. R., P. M. Edwards, M. J. Evans, J. D. Lee, M. D. Shaw, F. Squires and A. C. Lewis (2017) ‘Clustering approaches to improve the performance of low cost air pollution sensors’, *Faraday Discussions,* 200, pp. 621-637.

Stingone J. A., O. P. Pandey, L. Claudio and G. Pandey (2017) ‘Using machine learning to identify air pollution exposure profiles associated with early cognitive skills among U.S. children’, *Environmental Pollution,* 230, pp. 730-740.

Stirling, J. R., M. S. Zakynthinaki and B. Saltin (2005) 'A model of oxygen uptake kinetics in response to exercise: Including a means of calculating oxygen demand/deficit/debyt', *Bulletin of Mathematical Biology,* 67, pp. 989-1015.

Strava Metro (2015) 'Comprehensive User Guide: Version 2.0' (WWW) San Francisco: Strava Metro. Available at: http://ubdc.ac.uk/media/1323/stravametro_200_user_guide_withoutpics.pdf (Accessed 19 March 2018).

Svendby, S. R. (2016) *Mathematical Modelling of Heart Rate During Cycling Exercise*, Master of Science Thesis, University of Oslo. Available at: https://www.duo.uio.no/handle/10852/52702 (Accessed 14 November 2018).

Van Dijk J. and T. De Jong (2017) 'Post-processing GPS-tracks in reconstructing travelled routes in a GIS-environment: network subset selection in attribute adjustment', *Annals of GIS,* 23, 3, pp. 203-217.

Wallaart, J. C. (1997) 'A study of the relationship between heart rate and minute breathing volume at various levels of work demonstrating the spread between individuals in a group and the implications in industry, *The 1997 International Society of Respiratory Protection Conference, Amsterdam, Holland,* Mona Vale, Australia, pp. 1-17.

Wickler, S. J., D. F. Hoyt, E. A. Cogger and M. H Hirschbein (2000) 'Preferred speed and cost of transport: the effect of incline', *The Journal of Experimental Biology,* 203, pp. 2195-2000.

Winters, M., M. C. Friesen, M. Koehoorn and K. Teschke (2007) 'Utilitarian Bicycling: A Multilevel Analysis of Climate and Personal Influences', *American Journal of Preventive Medicine,* 32, 1, pp. 52-58.

Woodcock, J., M. Tainio, J. Cheshire, O. O'Brien and A. Goodman (2014a) 'Health effects of the London bicycle sharing system: health impact modelling study', *British Medical Journal,* 348, 7946, pp. 1-14.

Woodcock, J., M. Tainio, J. Cheshire, O. O'Brien and A. Goodman (2014b) 'Online Appendices', pp. 1-33. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3923979/.

Wright, M. A. E., L. Oppermann and M. Capra (2008) ‘Visualizing Data Gathered by Mobile Phones’, in Dodge, M., M. McDerby and M. Turner [eds] *Geographic Visualization: Concepts, Tools and Applications,* Chichester: Wiley, pp. 307-318.

Yin H. and O. Wolfson (2004) ‘A Weight-based Map Matching Method in Moving Objects Databases’, *Proceedings of the International Conference on Scientific and Statistical Database Management, SSDBM,* 16, pp. 437-438.

Zakynthinaki, M. S. (2015) ‘Modelling Heart Rate Kinetics’, *PLoS ONE,* 10, 4, pp. 1-26.

Zhao, S., X. Yuan, J. Zhang and Z. Li (2018) ‘AirNet: a machine learning dataset for air quality forecasting’, *ICLR 2018 Conference Blind Submission,* pp. 1-10.

Zhu, D., C. Cai, T. Yang and X. Zhou (2018) ‘A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization’, *Big Data and Cognitive Computing,* 2, 5, pp. 1-15.