Abstracts of Dissertations

Institute of Information and Communication Technologies

BULGARIAN ACADEMY OF SCIENCES

3 / 2025

1869

MODELING THE STATE OF AIR QUALITY BASED ON HEALTH AND ECONOMIC ASPECTS

Petar Zhivkov

МОДЕЛИРАНЕ НА КАЧЕСТВОТО НА АТМОСФЕРНИЯ ВЪЗДУХ ВЪРХУ ЗДРАВНИ И ИКОНОМИЧЕСКИ АСПЕКТИ

Петър Живков

Автореферати на дисертации

Институт по информационни и комуникационни технологии

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

ISSN: 1314-6351

Поредицата ***„Автореферати на дисертации на Института по информационни и комуникационни технологии при Българската академия на науките“*** *представя в електронен формат автореферати на дисертации за получаване на научната степен „Доктор на науките” или на образователната и научната степен „Доктор”, защитени в Института по информационни и комуникационни технологии при Българската академия на науките. Представените трудове отразяват нови научни и научно-приложни приноси в редица области на информационните и комуникационните технологии като Компютърни мрежи и архитектури, Паралелни алгоритми, Научни пресмятания, Лингвистично моделиране, Математически методи за обработка на сензорна информация, Информационни технологии в сигурността, Технологии за управление и обработка на знания, Грид-технологии и приложения, Оптимизация и вземане на решения, Обработка на сигнали и разпознаване на образи, Интелигентни системи, Информационни процеси и системи, Вградени интелигентни технологии, Йерархични системи, Комуникационни системи и услуги и др.*

The series ***Abstracts of Dissertations of the Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences*** *presents in an electronic format the abstracts of Doctor of Sciences and PhD dissertations defended in the Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences. The studies provide new original results in such areas of Information and Communication Technologies as Computer Networks and Architectures, Parallel Algorithms, Scientific Computations, Linguistic Modelling, Mathematical Methods for Sensor Data Processing, Information Technologies for Security, Technologies for Knowledge management and processing, Grid Technologies and Applications, Optimization and Decision Making, Signal Processing and Pattern Recognition, Information Processing and Systems, Intelligent Systems, Embedded Intelligent Technologies, Hierarchical Systems, Communication Systems and Services, etc.*

e-ISSN: 1314-6351

www.iict.bas.bg/dissertations

BULGARIAN ACADEMY OF SCIENCES

Abstract of PhD Thesis

MODELING THE STATE OF AIR QUALITY BASED ON HEALTH AND ECONOMIC ASPECTS

Petar Rumenov Zhivkov

Supervisor: Prof. Stefka Fidanova

Approved by Supervising Committee:

Prof. Ivan Dimov
Prof. Olympia Roeva
Assoc. Prof. Rayna Georgieva
Prof. Zlatinka Svetoslavova
Prof. Evdokia Sotirova

INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES
Department of Parallel Algorithms and Machine Learning with Neurotechnology Laboratory

Chapter 1

Introduction

Air pollution, particularly fine particulate matter (PM), is a leading environmental cause of morbidity and premature mortality, having risen from fifth position in 1990 [37]. Deaths attributed to ambient PM have increased over the past 25 years [13], and exposure to higher levels of atmospheric pollution is associated with increased mortality [7]. The latest report on the

PM is classified based on its size, with particles smaller than 10 micrometers (PM10) and 2.5 micrometers (PM2.5) causing the greatest concern due to their ability to penetrate deep into the lungs and cause adverse health effects. PM constitutes a complex mixture of solid particles and liquid droplets suspended in the air, with varying sizes and chemical compositions. These particles can originate from natural sources such as dust and pollen, as well as human activities including industrial emissions, vehicle exhaust, and combustion of biomass and solid fuels.

Studies have shown that exposure to PM2.5 and PM10 is associated with premature mortality from several diseases, including cardiovascular, respiratory, lung cancer, and upper and lower respiratory tract infections [2], [23], [26]. Urban air pollution increases the risk of pulmonary and systemic oxidative stress, hypoxemia, immunological modifications, atherosclerosis, and accelerated progression of cardiovascular diseases and chronic obstructive pulmonary disease (COPD) [32], while epidemiological studies indicate that PM exposure increases the risk of diabetes [60].

The severity of health consequences can vary depending on the concentration and duration of exposure, as well as the individual's health status and sensitivity to pollutants. Some population groups, such as children, the elderly, and those with pre-existing respiratory or cardiovascular diseases, are particularly vulnerable to the health effects of air pollution [45].

Studies show a strong association between air pollution and hospitalizations for respiratory diseases [52, 18, 33]. A cohort review of 175 articles summarizes that exposure to air pollution increases the likelihood of developing various diseases including respiratory diseases (asthma, chronic obstructive pulmonary disease (COPD), and lung cancer), cardiovascular diseases (heart attacks, strokes, and hypertension), kidney, liver, and others [1].

1.1 Relevance and Motivation of the Topic

The growing concern regarding the impact of air pollution on public health necessitates the development of predictive models to forecast acute morbidity levels based on air quality data [34]. The latest report by the European Union (EU) on air quality assesses the EU's efforts to improve air quality as only "partially effective" and specifically highlights that actions to reduce PM2.5 are insufficient. Figure 1.1 presents the results of a study involving 432 cities in the EU, which analyzes the economic consequences of the three main pollutants in European cities: particulate matter (PM), nitrogen dioxide, and ozone [15]. The results indicate that PM is the primary cause (82.5%) of economic damage resulting from illnesses and mortality due to air pollution. The study utilizes the AIRQ+ modeling program, developed by the World Health Organization (WHO), incorporating pollution levels and economic data for each city.

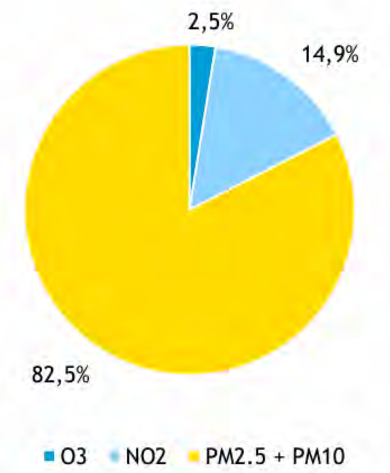


Figure 1.1: Contribution of the three main urban pollutants to the total morbidity and mortality damage in 432 European cities

In light of these findings, there is a growing need for models that can examine acute morbidity based on air quality data. Despite increasing awareness of the health risks associated with air pollution, there is a need for an in-depth analysis of specific pollutants and their effects on various health outcomes. Existing studies show links between air pollution and respiratory diseases, cardiovascular diseases, adverse birth outcomes, and even mental health issues. However, a comprehensive investigation of the extent and mechanisms of these connections is still required to inform evidence-based interventions and policies at the national and local levels.

The levels of PM exceed EU air quality standards in Bulgaria. The European Commission won a case against Bulgaria in April 2017 due to excessive air pollution in several cities. Bulgaria also leads in lost healthy life years due to dirty air. According to an international report, our citizens lose 2.5 years of their lives due to air pollution, whereas the EU average is 0.7 years. The main reasons for air quality surpassing the standards are household heating, car traffic, and geographical factors that trap pollutants in the atmosphere. The lack of solid research and statistical data on the health effects of PM pollution in Sofia undermines the city's ability to develop evidence-based strategies to mitigate its impact. Without accurate and detailed information, policymakers and public health officials face significant challenges in implementing targeted interventions and raising public awareness of the severity of the problem. Hence, there is an urgent need for comprehensive research efforts and improved data collection methods to overcome existing knowledge gaps and pave the way for informed decision-making in combating PM pollution and improving health quality in Sofia.

1.2 Aim and objectives of the thesis

The aim of the dissertation is to study the impact of fine particulate matter (PM) on acute illnesses in Sofia and to identify methods for prevention.

To achieve this aim, the following four tasks have been formulated:

- Investigate the relationship between fine particulate matter and health indicators for acute morbidity in Sofia;
- Improve data from citizen air quality monitoring stations through calibration using a machine learning-based two-step method;
- Develop a software tool for optimizing and evaluating cycling routes by characterizing cyclists' exposure to air pollution;
- Develop an IoT platform for aggregating and modeling air quality sensor data.

1.3 Research Methodology

The present dissertation employs a multidisciplinary approach, combining literature review, quantitative analysis, and advanced statistical modeling. A systematic review of existing studies has been conducted to summarize and synthesize the current evidence on the relationship between air quality and health outcomes. Additionally, epidemiological studies and data from health surveillance agencies have been analyzed to quantify the relationship between specific pollutants and health consequences. Modern statistical techniques, such as regression modeling and spatial analysis, have been utilized to investigate complex interactions and spatial patterns of air pollution and health outcomes.

The methodology used in this dissertation involves a systematic approach to studying the impact of fine particulate matter (PM) on acute diseases in Sofia and developing strategies to mitigate their consequences. The methodology includes data collection from hospitals and monitoring stations, data analysis to establish correlations, calibration of citizen laser stations, development of an IoT platform for data visualization, and optimization of cycling routes.

Data collection includes admissions and hospitalizations from two major hospitals in Sofia, emergency medical service registry data, and PM measurements from both official monitoring stations of the Executive Environmental Agency (EEA) and citizen laser stations. Meteorological data, including humidity, atmospheric pressure, and temperature, as well as traffic data and other relevant factors, are also collected.

Data analysis techniques are applied to investigate the relationship between PM levels and acute diseases. Statistical analysis and comparisons of PM levels below and above the World Health Organization (WHO) health standards are performed. The calibration process focuses on calibrating data from citizen laser stations using machine learning models with controlled and uncontrolled methods. Reference data from EEA monitoring stations are used for precise calibration, considering factors such as humidity, atmospheric pressure, and temperature.

A model for optimizing cycling routes to minimize PM inhalation is developed. This involves using a modified algorithm to find the shortest path and conducting real-field tests to validate the methodology. A software system is developed to visualize the location and indicators of monitoring stations on a geographical map. A new approach is applied for aggregating, organizing, processing, modeling, and exchanging data in the IoT system.

During the research process, a wide range of literature was utilized to investigate the problem. The newly developed methods and algorithms were published in [[**?**], [62], [63], [64]]. The presented methodology allows for a comprehensive analysis of the impact of fine particulate matter on acute illnesses, calibration of citizen laser stations through machine learning and neural networks, and optimization of cycling routes considering PM pollution. In line with the set objectives, algorithms and methodologies have been developed to address specific problems. Software applications for each of the developed algorithms have been created, and data visualization has been implemented through an IoT platform. The platform and software are written in Python and utilize the Django framework.

The results of this thesis will contribute to scientific knowledge, policy development and public awareness, ultimately leading to improved air quality management and better protection of human health.

Chapter 2

Link between PM Pollution and Acute Health Indicators

The links between urban air pollution (UAP) and human health have been consistently and clearly established by a number of researchers [53], [47], [31], with prominent groups of diseases being cardiovascular [55], neurovascular [51], [9], and pulmonary [54]. According to recent refined modeling, it is estimated that there are nearly 9 million deaths annually from UAP [44]. About 25% of premature deaths associated with UAP are respiratory by presumption [5].

Significant literature from epidemiological studies suggests a link between acute morbidity and exposure to PM2.5 air pollution [17]. Most of these data come from time-series analyses [20], comparing variations in hospitalizations with average variations in particulate matter [24]. UAP is responsible for respiratory tract inflammations [12], and considering that the respiratory system is a common gateway of entry, minimizing respiratory tract exposure also minimizes cardiovascular problems [49]. Additionally, there are studies that model pollution across multiple cities, such as the European Air Pollution and Health: A European Approach (APHEA) [29] and the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) [48], both providing consistent evidence of the health-air pollution link for multiple cities by covering large geographic areas.

The consequences of air pollution can be viewed as either an increased risk of illness or injury to an individual or as an additional general risk to the well-being of the population [19]. The goal of air quality management is to control or avoid the adverse effects of air pollution on public health. Therefore, it is important to define such impacts that are considered "adverse" and distinguish them from those impacts that are not considered adverse, thus concentrating protection efforts on pollutants that cause the most extreme health impacts.

2.1 PM Standards: Analysis of EU Legislation and WHO Guidelines

The EU legislative framework for PM levels in the air is based on the Ambient Air Quality Directive (2008/50/EC) and is complemented by the National Emission Ceilings Directive (2016/2284/EU). These regulations require Member States to meet air quality standards

Table 2.1: Air Pollution: WHO Recommendations and EU Legislation (in $\mu g/m^3$)

Air pollution standards	WHO		EU	
	PM10	PM2.5	PM10	PM2.5
daily average	50	25	50	–
yearly average	20	10	40	25

according to specific threshold values derived from scientific assessments. Compliance is monitored by the European Commission, and non-compliance can lead to legal actions and fines. In comparison, WHO provides general air quality guidelines without legislative obligations, leaving implementation to individual countries.

In the comparative analysis between EU legislation and WHO guidelines regarding PM2.5 levels in the air, several key differences emerge (Table 2.1). While both the EU and WHO recommend a daily average concentration threshold for PM10 of 50 $\mu g/m^3$, the EU and WHO significantly differ in their annual average recommendations, with WHO standards being twice as stringent (20 $\mu g/m^3$ compared to 40 $\mu g/m^3$ from the EU). For PM2.5, the EU does not have a daily average concentration threshold, and there are significant differences in annual recommendations, with WHO advocating for stricter limits. Furthermore, EU standards are region-specific, while WHO guidelines are non-binding and universal, providing guidance to countries worldwide.

For the purposes of this dissertation, we will use WHO air quality guidelines as they are recommended globally, have a daily limit for PM2.5, and are based solely on health aspects, compared to the EU directive, which focuses on regional and national needs and includes policy-economic and social aspects in the directives.

In its latest report, WHO pays special attention to PM pollution and considers exposure to PM2.5 as the greatest air-related killer globally. About 80% of global deaths attributable to PM2.5 exposure could be prevented if countries adhere to an annual PM2.5 threshold value of 5 $\mu g/m^3$. Achieving interim targets will also have significant health benefits. In the case of PM2.5, achieving interim target 4 (10 $\mu g/m^3$) would lead to approximately a 48% reduction in the total number of deaths attributable to PM2.5 exposure.

2.2 Geographical Study Area

Sofia is the only European capital located in a basin and is characterized by high levels of anthropogenic emissions and frequent occurrences of stagnant meteorological conditions. The city has a population of 1.2 million people [38] and is situated in the Sofia Basin. The area is recognized as problematic, particularly during winter when numerous exceedances of the norms in European legislation for air pollution control occur.

For years, Sofia has been struggling with significant levels of air pollution, with PM concentrations consistently exceeding the recommended limits set by WHO and the EU. These elevated levels of PM pollution in the city pose a serious threat to the health and well-being of its residents. Figure 2.2 shows the average concentration of PM2.5 in the most polluted capitals in Europe in 2022. It can be seen that Sofia ranks among the top positions in this ranking. Additionally, it is important to note that air quality measurement data are insufficient for Sofia. Data for PM2.5 levels in the capital are provided by only one station - AIS Hipodruma, which is insufficient for quality statistical analyses for a city the size and population of Sofia.

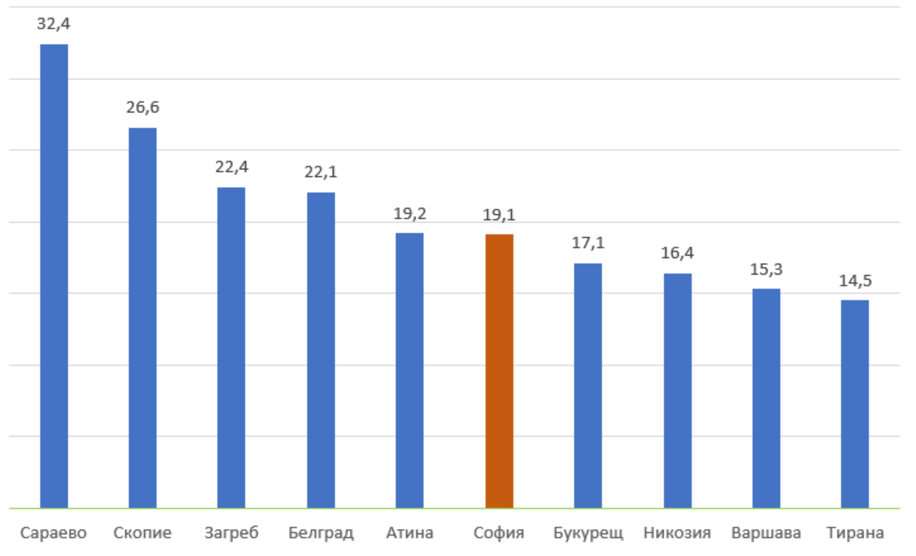


Figure 2.1: The average concentration of PM2.5 in the most polluted capitals in Europe in 2022 (in $\mu g/m^3$)

The combination of cold winters with subzero temperatures and the city's location predisposes it to temperature inversions that can last for several days to a week. Many urban areas situated in valleys with poor air exchange encounter significant air pollution problems, which are associated with local atmospheric conditions [46]. Strong inversions and a lack of precipitation and/or wind are the main reasons for retaining pollutants in the air, primarily during winter.

Historically, air quality indicators in Sofia during winter have been significantly worse than those during summer. In addition to temperature inversions typical of colder conditions, during the heating season, the burning of fossil fuels for domestic heating such as wood and coal, and sometimes waste, significantly increase PM levels.

Therefore, this difference in air quality between summer and winter conditions will serve as a good model for testing our hypothesis regarding a significant link between pollution and health outcomes in Sofia. The literature abounds with models that can be used to assess the hypothesis of pollution/health relationship. Specifically for Sofia, no such studies have been conducted. Correlation between temperature and mortality around Sofia is made using linear and nonlinear models [43]. Other studies examining the genotoxicity of atmospheric air in 3 European cities, including Sofia, show that air pollution in winter is 6 to 10 times higher compared to summer air [21].

2.3 Methodology of the Study

This is the first such study with real data from official and citizen sources of information on particulate matter air pollution, comprehensive data from the emergency medical service center (emergency aid), and data from two of the largest hospitals associated with access to emergency care in Sofia - UMBSM Pirogov and UMBSM Tokuda. The results can be compared and contrasted with other international studies using local data.

The aim of this study is to determine which acute morbidity increases and by how much on days when PM levels do not meet WHO guidelines, using time-series analysis.

The following four subtasks have been addressed: comparing air quality data with health data obtained from hospitals and emergency care; comparative analysis and highlighting of key problem areas; testing the hypothesis that data from inexpensive sensors can be useful in such epidemiological studies; identifying future prospects and summarizing key areas where additional research is needed to improve the model's effectiveness.

Air Pollution Data

First, we will identify which days, within the observed period, exceed the daily average limits set by the WHO for PM concentrations. Then, we will compare the health data for the days when air quality is within WHO guidelines to those that are not.

The air quality data in the form of hourly average concentrations of PM2.5 and PM10 are for the years 2018 and 2019 and represent independent series for each element from official sensors provided by (a) the Executive Environment Agency (EEA) and (b) laser stations from the luftdaten network (a global sensor network run by volunteers that generates open environmental data).

The data is collected from 5 official monitoring sites (Druzhba, Nadezhda, Hipodruma, Pavlovo, and Mladost). The method of beta attenuation is used for measuring PM10 and PM2.5 according to European Directive 2008/50/EC (Directive, 2008).

According to previous studies, the use of data from inexpensive laser stations aims to enhance the potential benefits of traditional monitoring networks with additional geographic and temporal resolution of measurements [8]. Additionally, inexpensive stations can complement official measuring devices and particularly address the limited data quantity and station distribution [65]. It is believed that inexpensive sensors are suitable for many specific purposes, including expanding conversations with communities and enhancing public awareness of air pollution issues [36]. In Sofia, at the time of the study, there are several citizen networks of laser stations that complement, with additional geographic and temporal resolution of measurements, the official stations of the EEA.

Health Data

Due to the voluntary nature of data provision, there are various temporal characteristics and formats from different sources:

- Aggregated data on the activity of the Sofia Emergency Medical Center by diagnoses from 01.01.2017 to 14.03.2019;
- Information on diagnostic activity from Tokuda Primary Care Center - from 02.01.2018 to 31.12.2018;
- Information on diagnostic activity from University Hospital "St. Marina" - from 01.01.2018 to 31.05.2019;
- Information on hospitalized diagnoses from St. Marina Hospital - from 01.01.2018 to 31.12.2018.

The study uses the International Classification of Diseases (ICD), specifically ICD-10, to segment diseases and identify morbidity. ICD-10 is the 10th revision of the International Statistical Classification of Diseases and Related Health Problems published by the World Health Organization (WHO), which is used up to the time of this study. It contains codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases [39], [40].

Methods

This study employs time series analysis with correlation methods to analyze air quality and health data. The statistical methods in the study fall into two categories: parametric and non-parametric. Parametric methods, such as Pearson's correlation coefficient, are preferred when the data are normally distributed or there is a sufficiently large volume of measurements, allowing for the assumption that the data follow a normal distribution. These methods are suitable for assessing the linear relationship between two continuous variables.

The parametric Pearson correlation test (2.1) is used to compare the two sources of air quality data. It provides a measure of the linear relationship between the two continuous variables (commonly referred to as just the correlation coefficient). Correlation coefficients for each pair (x, y) are calculated for assessment, and the values of x and y are replaced with their ranks, respectively. The results of the test are applied to the correlation coefficient in the range of -1 to 1.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2.1}$$

Non-parametric comparative methods, such as Spearman's rho and the Mann–Whitney U test, are used when the data are not normally distributed or there is uncertainty about whether they follow a normal distribution. These methods are suitable for assessing the rank correlation between variables or for comparing differences between two groups without requiring specific assessment of the linear relationship. The choice in this dissertation is as follows: descriptive analysis for individual variation series; correlation analysis (non-parametric - Spearman's rho) between pairs of variables; between-group comparison (non-parametric Mann–Whitney U test) of mean values.

The model developed in this study aims to explore the relationship between air quality indicators and health outcomes based on data on fine particulate matter and hospital admissions and Emergency Medical Service (EMS) visits. At the advice of medical professionals and health data analysts, we incorporate lagged effects of 1, 2, and 3 days. This is because health effects typically require time to manifest after exposure to air pollutants [28]. By doing so, our analysis of hospital admissions and EMS records will not only reveal which disease cases undergo changes due to variations in air quality but also demonstrate how these effects evolve over time. Figure **??** illustrates the exposure-response method, where the dose is the exposure multiplied by time.

2.4 Results

In a recent study, it is suggested that low-cost sensors can provide rough details about air quality but may not be suitable for applications requiring high accuracy [27]. In this chapter, we test the hypothesis that low-cost air quality sensors from Luftdaten can be a valuable addition to the official monitoring stations operated by the EEA. By cleaning and summarizing the data, we aim to identify correlations that can provide insights into certain coarse details. During comparison, we filtered out data when the air humidity exceeded 70%. While these conditions do not necessarily imply inaccurate data, the manufacturer does not guarantee the predicted accuracy of 10% under such circumstances.

The results of daily concentrations for PM10 and PM2.5 are summarized in Table ?? The average daily concentration of PM2.5 is 11.7 g/m3 with a range from 2 to 136 g/m3. During the study period, 13.2% of the daily concentrations of either PM10 or PM2.5 did not meet the WHO air quality guidelines (50 g/m3 for PM10 and 25 g/m3 for PM2.5). There were some gaps in the data from the EEA for several hours to several days (due to power supply issues, malfunctions, maintenance), and some of the sources for hospital admissions had missing data during weekends (Saturday, Sunday, official holidays). Therefore, an analysis of the gaps was conducted to exclude the possibility of bias.

Table 2.2: Daily concentration of PM10 and PM2.5, divided into working and non-working days

Day from the week	**N**					
PM10 Working days	522	34.17	27.5	5	336	20
PM10 Non-working days	230	37.1	26	6	290	18
PM2.5 Working days	522	11.15	9	2	104	4
PM2.5 Non-working days	230	12.93	8	2	113	3

The correlations observed in the data for hospitalized patients with more severe conditions are summarized in Tables 2.3 and 2.4. An increase (compared to background levels) in respiratory and pulmonary diseases is noted from the 1st to the 3rd day following an exceedance of PM10 by 120%, as well as heart failure during the same period by 19%. Exceedances of PM2.5 are associated with a 59% increase in the frequency of pulmonary embolism on days 2 and 3.

The results show an increase in emergency medical service contacts by 11% on days with exceedances of PM10 and 13.5% on days with exceedances of PM2.5, for a period of at least 3 days. The average increase (compared to background levels) in neuroses on day 2 following an exceedance of PM10 is 1%, for myocardial infarctions on day 3 following an exceedance it's 8%, for strokes immediately following an exceedance it's 9%, continuing until day 3 post-exceedance. The same applies to hypertensive conditions, which increase by 5% immediately around the exceedance and up to the third day after the event.

From the analyzed data, a significant correlation is observed between cardiovascular and cerebrovascular diseases, fully coinciding with the rapid increase in risk at low levels

Table 2.3: Comparison of nonparametric data from hospitalized patients on days with exceedances of WHO standards for PM10

Morbidity	ICD-10	no lag		1-day lag		2-days lag		3-days lag	
		Z	p	Z	p	Z	p	Z	p
Resp. system	C30-C39	-1,504	0,133	-2,211	0,027	-2,870	0,004	-2,862	0,004
Heart attack	I50	-1,729	0,084	-3,656	0,001	-3,475	0,001	-2,821	0,005

Table 2.4: Comparison of non-parametric data from hospitalized patients in severe conditions on days with exceedances of WHO standards for PM2.5

Morbidity	ICD-10	no lag		1-day lag		2-days lag		3-days lag	
		Z	p	Z	p	Z	p	Z	p
Resp. system	C30-C39	-0,361	0,718	-1,270	0,204	-1,265	0,206	-2,119	0,034
Pulm. embolism	I26	-1,601	0,109	-2,302	0,021	-2,894	0,004	-1,852	0,064
Heart attack	I50	-1,586	0,113	-2,831	0,005	-3,195	0,001	-2,799	0,005

Table 2.5: Comparison of non-parametric data from outpatient examinations in hospitals with WHO standards for mean daily concentrations of PM10

Morbidity	ICD-10	no lag		1-day lag		2-days lag		3-days lag	
		Z	p	Z	p	Z	p	Z	p
Respiratory system	C30-C39	−1.515	0.130	−2.950	0.003	−2.941	0.003	−2.423	0.015
Ear diseases	H65,H81	−0.754	0.451	−0.976	0.329	−1.580	0.114	−2.322	0.020
Myocardial infarct.	I20-I25	−0.244	0.807	−1.393	0.164	−1.876	0.061	−2.104	0.035
Heart deficiency	I50	−0.771	0.440	−2.840	0.005	−2.559	0.010	−2.972	0.003
Upper respir. tract	J00-J06	−3.218	0.001	−3.507	0.000	−3.910	0.000	−4.605	0.000
Bacterial pneum.	J13-J18	−3.171	0.002	−3.395	0.001	−2.556	0.011	−3.683	0.000
Acute bronchiolitis	J20-J21	−3.096	0.002	−2.848	0.004	−1.606	0.108	−1.588	0.112
Lower resp. tract	J40-J47	−3.988	0.000	−4.049	0.000	−2.842	0.004	−3.481	0.001
Skin infections	L00-L08	−2.552	0.011	−1.379	0.168	−0.267	0.789	−0.489	0.625
Mycoplasma pneum.	J20.0	−2.330	0.020	−2.327	0.020	−0.430	0.667	−0.431	0.667
Acute bronchitis	J20.9	−2.496	0.013	−1.449	0.147	−0.039	0.969	−1.292	0.197
Asthma, allergic	J45.0	−2.426	0.015	−2.808	0.005	−3.577	0.000	−3.643	0.000
Angina	I20.8	−1.971	0.049	−0.763	0.445	−0.773	0.440	−0.674	0.500

of PM2.5 pollution fractions, consistent with similar findings in other studies [61], [50], [56].

But the highest correlation is observed in acute upper respiratory tract infections, increasing by 47%, and notably pneumonia by 60%. Chronic obstructive pulmonary disease (COPD) increases by 36% after one day. Regarding asthma, records of allergic asthma increase more on days with increased pollution compared to non-allergic asthma. Ambulatory hospital visits for mild conditions that do not require hospitalization logically show an increased presence of respiratory diagnoses (Table 2.5).

2.5 Conclusion

The analyses and tests conducted have proven that acute illness increases on days when the concentrations of fine particulate matter exceed the WHO air quality standards. This was established by comparing hospital admissions and emergency cases on days with and without exceedances of PM standards in Sofia.

Respiratory diseases show higher and faster development when air pollution exceeds healthy limits. Other diseases related to air pollution, such as cardiovascular and cerebrovascular diseases, also increase.

This is among the first studies of its kind in Bulgaria and could be valuable for healthcare professionals, environmental scientists, and policymakers.

The exposure-effect method is complex, and achieving precise measurements is challenging, even in exposure chambers. However, with the help of epidemiological research methods and population statistics, it becomes evident that air pollution contributes to expected changes in certain disease indicators. In conclusion, similar to data from published studies in cities in Europe, America, and Asia, elevated levels of air pollution are associated with and lead to higher levels of diagnosed diseases. With few exceptions, short-term health effects are measured using averaged concentrations of air pollution across the entire city as exposure indicators. This can lead to misclassification of exposure and thus bias.

The use of extensive data from multiple low-cost laser sensors can effectively enrich the information provided by the high-precision official stations of the Executive Environment Agency, which, besides being few in number, often have missing data for certain periods. This approach leads to improved measurements of the links between health and air pollution by providing a more comprehensive exposure to pollutants.

Chapter 3

Development of a software model for bicycle route selection with respect to PM pollution

Most of the specialized and segregated bicycle lanes in cities are located near busy traffic arteries, which can pose a significant health risk to cyclists due to increased pollutant intake associated with higher ventilation rates and physical activity levels [66], [22], and high levels of physical activity [6], [42]. Researchers have focused on assessing actual exposure levels of cyclists along predetermined routes using personal samplers [14] or deriving personal exposure from street-level pollution measurements using mobile laboratories or stationary air monitoring stations [41]. Many studies also attempt to link specific physiological reactions to cyclists' exposure to air pollution and find evidence that short-term exposure can lead to adverse health effects [57]. One study even found that cyclists absorb a greater portion of fine particles PM2.5 and black carbon than drivers of motorized transport [25].

3.1 Formulation of the Problem

Here, the issue of the lack of quantitative determination of exposure of urban cyclists to PM in areas with poor air quality is addressed. In the city of Sofia, known for its high levels of air pollution, there is a lack of a comprehensive method for assessing and quantitatively determining the exposure levels of cyclists. This knowledge gap hinders the development of effective strategies to minimize exposure and provide healthier cycling routes.

To address this problem, we propose the development and evaluation of a software model that utilizes a modified route determination method reflecting the expected inhalation of PM by the cyclist, providing a healthier route. By incorporating data on air quality, pollutant concentrations, and traffic patterns into the software model, we can quantify cyclists' exposure to PM and provide recommendations for routes with minimal exposure. This innovative approach aims to fill the existing gap in quantitative exposure determination and provide a practical tool for urban cyclists in polluted cities like Sofia.

The software implements the decision-making system outlined in this chapter. It is deployed in a cloud infrastructure, and Python, Django, and GraphQL (for the database) were used for its development. It combines GIS data with traffic and air quality data to apply the modified $k - SPwLO$ algorithm to find the most suitable bicycle route and display it on the OpenStreet map.

The input parameters include: PM measurements from fixed sensors; PM measurements from mobile sensors; traffic data; temperature and humidity from stationary sensors; start and end points of the journey; terrain elevation; heart rate.

3.2 Research study

This study examines the creation of a software tool aimed at selecting the optimal cycling route that minimizes the inhalation dose of PM for a cyclist traveling from point A to point B.

The inhalation dose (ID) of air pollutants is a variable that depends on pollutant concentrations, time, and ventilation rate (VR) [min]. We calculate the inhalation dose by incorporating PM exposure for each cyclist with biomarkers such as heart rate and the time required to travel each route into the model. The next subsection will provide more detailed information on the calculation methods used in the model.

To calculate the cyclist's ventilation rate (VR) (in l/min), we use the equation from the model [16], which is based on the heart rate (HR) [min] (Equation 3.1).

$$VR = 0,00070724 \times HR^{2,17008537} \tag{3.1}$$

To determine the quantities of particulate matter affecting cyclists, we use Equation 3.2 [16] to calculate the inhalation dose of PM for each segment:

$$_{inh} = \mathrm{PM}_{conc} \times VR \times time \tag{3.2}$$

where PM_{inh} (μg) is the mass of pollutants that enter the cyclists' respiratory tract during travel (in both directions); PM_{conc} ($\mu g/m^3$) is the average exposure to pollutants.

The formulas for calculating the inhalation dose (Equations 3.1 and 3.2) lead to the following hypothesis: If we want to create a tool that reduces the inhalation dose of PM, it should select a route that is fast and short. The shorter the time, the smaller the PM_{inh}; it requires less effort. HR increases during uphill climbs and high speeds. Look for routes with low elevation gain. Pass through areas with lower concentrations of PM; avoid heavily trafficked roads. Prefer smaller streets and park alleys.

The decision-making system is calibrated using data on HR, IR, cycling speed, PM measurements from mobile and stationary sensors, and traffic data. The software integrates traffic and air quality data from multiple sources with specific bicycle route data, including exact location, duration of ride, and elevation gain.

To calculate the least inhalation of polluted air, this software first needs to compute the shortest path in linear time [4], [58]. Additionally, it should present options that are longer than the shortest path but have different desired characteristics, such as lower vehicle traffic and minimal elevation gain. The k-Shortest Paths problem is a simple method for computing alternative routes [35].

Therefore, in this study, we decided to use an alternative routing, specifically k-Shortest Paths with Limited Overlap (k-SPwLO), which was introduced earlier in [10].

The k-SPwLO algorithm searches for paths that are: as short as possible; sufficiently different from each other.

Despite the method's better performance compared to the baseline solution, which lists paths in increasing order of their length, OnePass is not useful even for moderately sized road networks [11]. For this purpose, we use MultiPass - a more precise method that, by adding a second pruning criterion, extends and improves OnePass. MultiPass traverses the network k-1 times, but evaluates and expands only the most promising paths, unlike OnePass, which traverses the road network once and expands only those paths that meet a similarity criterion. Pruning is performed for each path that cannot lead to a solution.

Let P be a set of routes in the road network G connecting nodes s and t. A path p' (s → t) is called an "alternative" in P when p' is sufficiently different from every path p P. Formally, the overlap coefficient between p' and p determines how similar they are:

$$Sim(p', p) = \frac{\sum_{(n_x,n_y)\in p'\cap p} w_{xy}}{l(p)}, \tag{3.3}$$

where $p' \cap p$ denotes the set of edges that are common to p' and p. Given the similarity threshold θ, the route p' is an alternative to the set P if $Sim(p', p) \leq \theta$.

Given a source node s and a target node t, the k-SPwLO query provides a collection of k routes from s to t, arranged in increasing order of length, such that: the shortest route $p'(s \to t)$ is always included; all k routes are pairwise dissimilar with respect to the similarity threshold θ; all k routes are as short as possible.

The final result of the `k-SPwLO-modified` function is a single path. This path represents the shortest path from the source node to the target node, taking into account certain constraints such as avoiding major boulevards and locations with high levels of $PM_{2.5}$ or heavy traffic. The algorithm traverses through the neighbors of each node, considering their relevant characteristics based on the specified criteria, and selects the path with the lowest cost. If the algorithm successfully reaches the target node, it returns this single path.

3.2.1 Measurements from fixed sensors

The software utilizes an aggregation tool that extracts air quality data from both standard environmental monitoring instruments by IAOS and inexpensive sensor networks such as Luftdaten, Smog, Openaqi, and many others. It receives, records, cleans, and calibrates air quality data from fixed inexpensive sensors.

To mitigate the aforementioned limitations of fixed inexpensive sensors, we calibrate the data obtained from them. This is done by examining data from both environmental monitoring stations and the Luftdaten network of inexpensive sensors. To calibrate the data from the inexpensive sensors and enhance their reliability, we employ a two-step calibration method [62], which utilizes artificial neural networks and anomaly detection.

3.2.2 Measurement of PM exposure using mobile sensors

Portable low-cost pollution sensing devices are environmental monitoring devices that people can carry or wear while going about their daily activities. Since they detect pollution levels directly and in real-time, they can enable healthcare providers and researchers to monitor exposure at an individual level and empower citizens to manage their personal pollution exposure beyond what regulatory monitoring stations can achieve [59].

The particle sensor in the Airbeam2 - PMS 7003 is the seventh generation in the PMSx003 series developed by Plantower. The PMS7003 is a digital and universal particle concentration sensor used to measure the quantity of suspended particles in the air and output the results through a digital interface. It can be used with various instruments to provide accurate real-time concentration data. It provides a solid particle counting efficiency of 98% and can operate within a temperature range of -10 to 60 degrees Celsius.

3.2.3 PM Measurements and Calibration of the Decision-Making Model

As previously discussed, the software is programmed to search for a route where: the shortest path $p'(s \to t)$ is always included; the routes are pairwise dissimilar with respect to the similarity threshold θ; the routes are as short as possible.

The computational model relies on the previously mentioned formulas for calculating ventilation and inhalation rates, which are used to develop the weights for the search algorithm.

Additionally, the algorithm is further enhanced through route assessment, combining traffic data with air quality data from mobile and stationary sensors. The integration aids in more accurately forecasting the computation of the optimal route. Stationary sensors positioned near bicycle routes establish the baseline concentrations of PM pollutants in the area - $PM_{baseline}$. Meanwhile, traffic data provides information to estimate the PM concentrations from vehicles and transportation - $PM_{traffic}$. The expected exposure to PM pollutants PM_{exp} along the bicycle route is determined by combining these two variables.

$$PM_{exp} = PM_{baseline} + PM_{trafic} \tag{3.4}$$

As the bicycle route approaches main roads, the influence of traffic pollution on overall PM concentrations increases. Therefore, the variable $PM_{traffic}$ increases. The time of travel also affects this variable - passing during rush hours and near heavily congested streets and boulevards increases PM concentrations, as noted further in the chapter. Actual PM concentrations are estimated using measurements from mobile sensors. Mobile sensors provide highly accurate measurements as they are attached to the bicycle itself, transmit PM measurements every minute via a mobile phone, and reflect ultra-local pollution, such as being directly behind a bus or large truck. Later in the field studies, this assessment process is demonstrated.

The characteristics for evaluating bicycle routes are highly dependent on the location and features of the specific area. This is why we conducted a very detailed field study, examining two bicycle routes described in the next section.

3.2.4 Field Tests and Investigated Routes

In this section, the software model is evaluated through real-world field testing in Sofia, examining two bicycle routes. Both investigated routes start from the Vasil Levski National Stadium and end at "Petite Kiosque." These two locations are situated in the center of Sofia and experience active bicycle traffic.

The two selected routes were chosen based on the following criteria: they have the same starting and ending points; one route is suggested by the software while the other follows dedicated bicycle lanes; both routes are actively used by cyclists; both routes pass near stationary sensors from the Luftdaten network; and both routes are deemed safe for regular cyclists.

On the following images, you can see the two paths used for evaluation: Segment A (Fig.3.1), which is the optimized bicycle route proposed by the software tool, passing through the shortest path via small central streets and parks; while Segment B (Fig.3.2) utilizes the developed bicycle infrastructure (mainly cycling near major road arteries with heavy traffic), is longer in distance, and is proposed by the navigation software as it follows dedicated bicycle lanes.

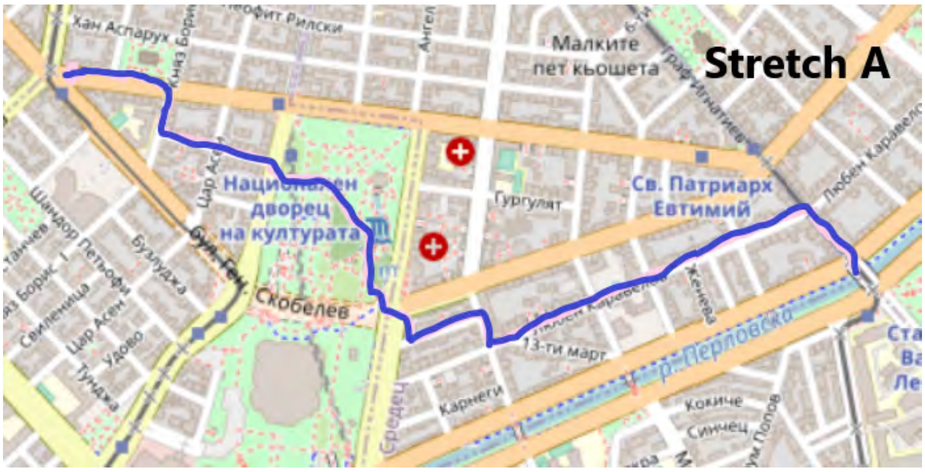


Figure 3.1: Section A - suggested by our software looking for the smallest inhalation dose

Figure 3.2: Section B - offered by most navigation software, which follows a designated cycle lane

Field tests were conducted with 10 participants in the study. The mobile equipment was attached to the front part of each bicycle, allowing the sampling lines to capture pollutants without obstruction; it was also secured from below to reduce vibrations.

We selected 10 individuals (aged 27-41) through contact with the local cycling community (8 males and 2 females). Before participating in the study, participants completed a pre-screening questionnaire. Exclusion criteria from the group included respiratory (including asthma), cardiovascular, or other chronic diseases, as well as tobacco smoking (current or recent). Only individuals who commuted daily by bicycle in Sofia were included. These factors were used to minimize the risk of injury due to unfamiliarity with Sofia's streets, lack of cycling experience, and adverse health consequences. Additionally, participants were asked to abstain from alcohol and caffeine for 48 hours prior to the tests.

Circular journeys were conducted on weekdays during rush hours (HT) (8:00-9:30 a.m.) and during off-peak hours (LT) (10:30 a.m.-12:00 p.m.) and on non-working days (NWD): weekends and holidays. The results from the three scenarios are evaluated separately in the next section. Field tests were conducted in April and May when the weather was moderate, and air pollution from domestic heating did not influence the data. Therefore, increases in particulate matter concentrations were mainly due to traffic. The ten participants performed circular journeys on both segments during HT, LT, and NWD.

3.3 Results and discussion

3.3.1 PM1 and PM2.5 concentrations from mobile measurements

In Table 3.1, the minimum, maximum, and mean (median) values of PM10 and PM2.5 concentrations for the two surveyed segments during working days under heavy traffic (HT) are shown. The concentrations of ultrafine particles with a diameter below 1 micron (PM10) ranged between 8 and 24 μg/m3 (average 11 μg/m3) for segment A and between 8 and 41 μg/m3 (average 14 μg/m3) for segment B. The concentrations of fine particles below 2.5 microns (PM2.5) ranged between 12 and 29 μg/m3 (average 15 μg/m3) for segment A and between 12 and 45 μg/m3 (average 19 μg/m3) for segment B.

Table 3.1: PM1 and PM2.5 concentrations from mobile measurements on weekdays during heavy traffic (HT)

		1			2,5	
	min	max	mean	min	max	mean
Stretch	8	24	11	12	29	15
Stretch B	8	41	14	12	45	19

In Table 3.2, the concentrations of PM10 and PM2.5 for segments A and B during working days under light traffic (LT) are presented. Measurements from segment A show lower concentration levels for both PM10 and PM2.5. The concentrations of ultrafine particles with a diameter below 1 micron (PM10) ranged between 6 and 19 μg/m3 (average 10 μg/m3) for segment A and between 5 and 34 μg/m3 (average 12 μg/m3) for segment B. The concentrations of fine particles below 2.5 microns (PM2.5) ranged between 10 and 24 μg/m3 (average 13 μg/m3) for segment A and between 9 and 38 μg/m3 (average 15 μg/m3) for segment B.

Table 3.3 illustrates the minimum, maximum, and median values of PM10 and PM2.5 concentrations for both surveyed segments during weekends and holidays.

Table 3.2: Concentrations of PM1 and PM2.5 from mobile measurements on weekdays with low traffic (LT)

		1			2,5	
	min	max	mean	min	max	mean
Stretch	6	19	10	10	24	13
Stretch B	5	34	12	9	38	15

Measurements for segment A show almost identical low concentration levels for PM10 and PM2.5. PM10 concentrations range between 3 and 7 μg/m3 (average 4 μg/m3) for segment A and between 3 and 8 μg/m3 (average 4 μg/m3) for segment B. PM2.5 concentrations range between 4 and 11 μg/m3 (average 7 μg/m3) for segment A and between 5 and 11 μg/m3 (average 7 μg/m3) for segment B.

Table 3.3: Concentrations of PM1 and PM2.5 from mobile measurements on non-working days

		1			2,5	
	min	max	mean	min	max	mean
Stretch	3	7	4	4	11	7
Stretch B	3	8	4	5	11	7

Despite the fact that the dedicated bicycle route is mostly exposed, the high volume of cars, buses, and trucks in this corridor is the main reason for the elevated pollution concentrations, as increased traffic raises the levels of PM. As a result, PM2.5 measurements are almost identical during weekends but with nearly 20% higher concentrations between the two routes during weekdays.

3.3.2 Ventilation Rate

To measure the ventilation rate (VR) during breathing, we conducted measurements of heart rate (HR), oxygen saturation (SpO2), and respiratory rate (RR) for each participant in the project. The average value of each parameter has low variability, and the final results for VR are presented in Table 3.4. No differences are observed in these three factors depending on whether the route is traversed on a workday or a non-workday, as they are not directly influenced by traffic or different levels of short-term air pollution exposure.

Table 3.4: Averaged indicators from field measurements

	Stretch	Stretch B
VR (L/min)	10,14	11,06
HR (beats/min)	82,30	85,66
Denivelation (m)	3	15
Speed (km/h)	14,3	14,7

Despite having the same starting and ending points, the distances are different, with

Segment A being shorter - 3.8 km compared to 4.4 km for Segment B, and it is faster - 15:52 minutes compared to 18:04 minutes for Segment B. The time taken to traverse both segments remains relatively constant on workdays and non-workdays. The specialized bike route had longer and straighter corridors with fewer intersections and crossings, leading to higher maximum and average speeds. Additionally, Segment B has a greater elevation change - 15 m compared to 3 m in Segment A. All these data result in increased cyclist speeds in Segment B - 11.06 km/h compared to 10.14 km/h for Segment A.

Table 3.5 shows the results of the inhalation dose of cyclists for PM10 and PM2.5. They are calculated based on the average exposure value of PM during the ride measured by mobile sensors, along with the time taken for the ride and the ventilation rate for each cyclist.

Table 3.5: Inhalation Dose of PM10 and PM2.5 during the Circuits

Period	Pollutant	Stretch A(optimized)	Stretch B(bike lanes)
Heavy traffic	PM 1_{inh}	29.74	46.45 (↑56%)
	PM2.5_{inh}	40.56	63.04 (↑55%)
Low traffic	PM 1_{inh}	27,04	39,82 (↑47%)
	PM2.5_{inh}	35.15	49.77 (↑42%)
Non-working days	PM1_{inh}	10.82	13.27 (↑23%)
	PM2.5_{inh}	18.93	23.23 (↑23%)

Despite having the same starting and ending points, Segment B shows an increased inhalation dose of particulate matter (PM) due to the longer duration of the route, requiring more effort, and exposing the cyclist to higher concentrations of PM due to vehicle exhaust gases. Even on non-workdays, when PM concentrations are similar across both routes, Segment B exhibits higher inhalation doses due to the longer duration and higher ventilation rate (due to higher elevation change and higher average speed).

Measurements for SpO2 and RR do not show any significant short-term health effects. This was expected since the study participants, selected for safety reasons, were non-smokers without chronic conditions and were regular cyclists. However, this does not necessarily mean that people with chronic conditions and sensitivity to air pollution may not experience symptoms or irritations, as observed in some studies.

3.3.3 Visualization of Bicycle Routes by Incorporating Data from Stationary Sensors

We will utilize the network of inexpensive stationary sensors from Luftdaten to demonstrate the technology for our fixed sensors in our software. Figure 3.3 displays the concentrations for Segment B during LT, where measurements from stationary sensors within 200 m or closer to the route are applied. The black-colored line indicates that in this part of the route, there is no fixed sensor closer than 200 m, while the green and yellow lines represent the concentrations measured by the fixed sensors nearby. Green color is used for PM2.5 concentration values between 0 and 12, while yellow indicates PM2.5 values between 12 and 35. These color categories are inspired by the EPA Air Quality Index and are the same for visualizing Aircasting routes measured with portable sensors. Thanks to the dense network of inexpensive sensors in Sofia, 5 fixed sensors are used for Segment A, and 6 sensors are used for Segment B, as they meet the selection criteria.

The software has continuous access to particulate matter (PM) concentrations from fixed sensors, while traffic data can dynamically change the input values for the decision-

Figure 3.3: Biking Route with Exposure from Fixed Sensors

making mechanism. The tool also has the functionality to display the same route in different time frames and accordingly show different air pollution concentrations.

3.3.4 Refinement of the model and evaluation of data from mobile and stationary sensors

The average values for PM2.5 concentrations from stationary sensors, located near the surveyed routes (within 200 meters), are compared with PM2.5 measurements from mobile sensors, as shown in Table 3.6 below. We find a strong correlation between the average values of $PM2.5_{fixed}$ and the minimum values for $PM2.5_{mobile}$, partly proving the assumption that the concentration from nearby fixed sensors can be used as an output for calculating the overall PM concentration estimate.

Data from the fixed sensors closely match the minimum concentrations along the route. Typically, this part of the route has negligible traffic, and its value often approximates $PM_{baseline}$, measured by the fixed sensors. Meanwhile, the maximum PM concentrations along the bicycle route are observed when passing by significant sources of pollution such as trucks, vehicles with old diesel engines, mopeds, or queues of cars waiting at traffic lights. As expected, the difference between the minimum and maximum PM concentrations is more significant during HT.

3.4 Conclusion

This chapter presents the development of a software tool that optimizes bicycle routes based on algorithms predicting the least harmful air pollutants. The algorithm utilizes a modified implementation of the k-shortest path with limited overlap method. It relies on experimental data and equations that calculate the total inhaled pollutant dose for cyclists. Additionally, the study evaluates two bicycle routes: Route A, suggested by the newly developed software, which traverses small streets in Sofia, and Route B, suggested by navigation apps, which follows a dedicated bicycle lane.

Table 3.6: PM2.5 concentrations from fixed sensors near the route and mobile measurements

		$2,5_{fixed}$		$2,5_{mobile}$	
Period	Route	mean	min	max	mean
WD-HT	Stretch A	10	12	29	15
	Stretch B	10	12	45	19
WD-LT	Stretch A	8	10	24	13
	Stretch B	8	9	38	15
NWD	Stretch A	4	4	11	7
	Stretch B	4	5	11	7

In the field test, ten cyclists make circular trips on both routes over three periods: (1) during heavy traffic, (2) during light traffic on workdays, and (3) on non-workdays. Based on the data collected during the study from a mobile sensor, cyclists' exposure and potential inhalation doses of PM2.5 and PM10 along both routes are calculated.

Exposure concentrations along dedicated bicycle lanes were found to be higher than exposure levels along the optimal route, especially on workdays. Even when average concentrations were nearly equal, the inhalation dose for cyclists was always higher on the bicycle lane route, as it took longer in terms of time and distance, had higher elevation differences, and required more intense cycling. Choosing the optimized bicycle route reduced the inhalation dose of PM2.5 by 23% on non-workdays and up to 55% during heavy traffic on workdays. Results during workdays show an additional health risk for cyclists using the investigated bicycle lanes due to pollutants directly associated with traffic. Exposure to PM concentrations in the investigated bicycle lanes was closely linked to vehicular traffic, as the study was not conducted during the heating season.

The findings of this study build upon previous observations [3], which suggest that redesigning streets for low-speed multimodal traffic without barriers is a more sustainable and pragmatic approach than constructing bicycle infrastructure on heavily trafficked roadways. There is a significant difference between studies conducted during light traffic and heavy traffic conditions. Bicycle lanes without physical barriers between the bicycle route and the road have higher exposure levels, a conclusion also noted in another study [30]. Our findings contribute to a better understanding of Sofia's air pollution issues, particularly related to vehicular traffic.

Chapter 4

Software System for Visualization of Air Pollution

In order to utilize all the research described up to this point and to convey information on how air pollution affects health in an accessible way, it is necessary to develop a comprehensive software system that allows for the collection, processing, and modeling of data.

The architecture of the software is modular, defining how different components interact, how data is processed, and how results are visualized. Figure 4.1 illustrates its appearance. The software architecture consists of the following components: API connection module; Central cloud repository; Data processing module; Forecasting module; Module for measuring PM concentration along a specified route; Data visualization and dissemination.

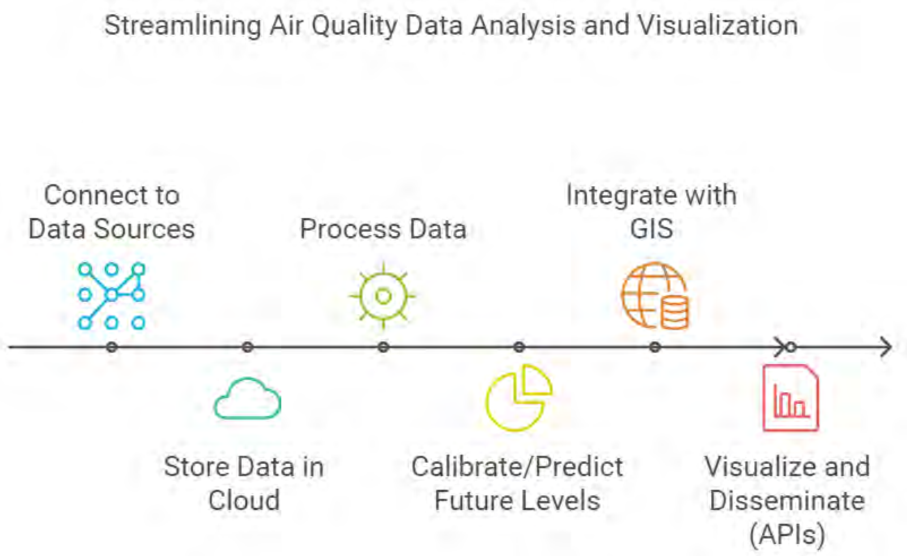


Figure 4.1: Software architecture

Ensuring the accuracy and reliability of our decision-making system is paramount to the success of our research. To achieve this, our software relies on a series of input parameters. These parameters include: measurements of PM concentration from stationary laser sensors; measurements of PM concentration from official monitoring stations; measurements of PM concentration from portable sensors; data on temperature and humidity from stationary sensors; traffic data.

The inclusion of these parameters is contingent upon their significance for assessing air quality and their potential impact on human health. Each parameter contributes to a comprehensive understanding of air quality, enabling our decision-making system to make informed forecasts and assessments.

To achieve this, we are committed to creating a modern air quality monitoring system that collects data from various sources.

The output data include: real-time air quality data by location, geographic radius, or city; station locations with their values on geographic maps; air quality forecast; calibrated data from specific stations; selection of the most suitable route based on PM pollution and traffic.

4.1 Data Collection Module

Access to data from national and civilian air quality monitoring networks is facilitated through public APIs (such as Luftdaten, aqicn, openaq, etc.). The built system periodically connects to these APIs over a specified period. These data provide us with a comprehensive view of air quality in a given area.

- Air Pollutants: We receive information about various air pollutants, including particulate matter (PM2.5 and PM10), nitrogen dioxide (NO2), sulfur dioxide (SO2), and carbon monoxide (CO). Each pollutant contributes to the overall air quality index.

- Traffic Data: Real-time updates on road traffic help us understand how vehicle activity affects air quality. This data source helps us identify congestion points and emissions.

- Temperature: Knowledge of temperature is essential as it affects the dispersion and concentration of pollutants in the air. Higher pollution levels are often observed during the winter heating season and temperature inversions.

- Humidity: Humidity levels influence air quality by affecting the concentration and dispersion of particles and pollutants in the air. High humidity can improve air quality, while low humidity can exacerbate issues. Humidity is also an important factor when comparing and processing pollution data from different measurement devices. Laser measuring stations show higher error levels when humidity is high.

- Air Pressure: Air pressure data helps in forecasting meteorological conditions that may impact air quality. This is an important factor in understanding atmospheric stability and modeling air pollution forecasts.

4.2 Cloud Data Storage Module

The system for monitoring air quality and its impact on human health is based on an integrated approach that involves collecting and processing diverse data. We gather data on air pollutants, road traffic, temperature, humidity, and pressure from various sources and centralize and store them in the cloud infrastructure. This approach ensures scalability and flexibility, allowing us to meet the growing demands of the system.

By collecting data on air pollutants, road traffic, temperature, humidity, and pressure, we can provide a comprehensive view of air quality. This approach not only enhances our understanding of air quality but also enables us to make informed decisions and take timely actions to protect public health and the environment. Traffic data is integrated as it provides information on vehicle emissions and their impact on air quality. By combining traffic data with air quality measurements, we can identify areas with high pollution due to congestion and develop strategies to reduce pollution.

The data is updated in real-time, allowing us to provide users with up-to-date information on air quality. Integrating data from various sources makes analysis and information extraction easier and more efficient, providing a unified platform for data processing. Users with different privileges have easy access to information from anywhere with an internet connection.

4.3 Module Analysis, Forecasting, and Air Pollution Measurement

The Data Processing Module is the heart of the system and performs several key functions. Figure 4.2 provides a comprehensive architectural overview of the air quality monitoring system. It visually outlines the complex data flow between the data processing, forecasting, and pollution measurement modules. The diagram clearly demonstrates how data is processed, used for forecasting, and measured in real-time. Additionally, it emphasizes the integration of geographic context through GIS data, facilitating a holistic understanding of the system's functionalities.

First and foremost, this module provides basic data preprocessing, including cleaning and structuring. This means that data obtained from various sources and APIs undergo analysis and processing to ensure their accuracy and integrity. Cleaning involves removing invalid or corrupted data, while structuring them in a suitable format enables easier navigation and analysis of the information. It is important to note that the module includes a two-step machine learning-based calibration method, which enhances the accuracy of the data and prepares it for analysis.

4.4 Used technologies

In this section, we will focus on the reason for choosing Python and the related libraries for data modeling within our software system for measuring air quality.

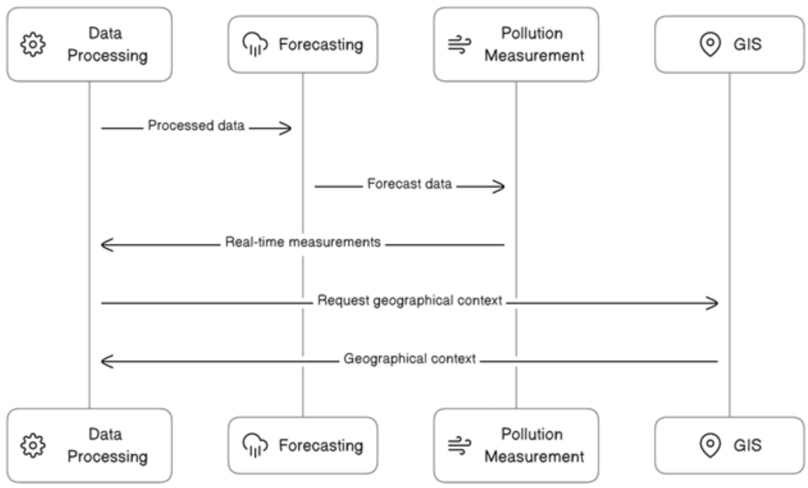


Figure 4.2: Architectural diagram showing the flow of data between the data processing modules.

We will provide information on how Python and these libraries integrate with the system and what key functionalities they provide.

Python is used as the primary programming language for developing the system, providing flexibility and power in data processing. Its readability and rich set of libraries make it a suitable choice for data processing and analysis.

Using Python and modeling libraries brings several significant advantages to our software system. Python provides a convenient environment for developing and testing models, allowing for rapid prototyping and efficient data modeling.

Python provides an extensive and flexible ecosystem of libraries that are crucial tools for data processing, including data obtained from financial and physical measurements. These libraries, such as NumPy, pandas, and SciPy, offer powerful tools for analysis, manipulation, and processing of numerical data. For instance, NumPy provides capabilities for efficient operations on multidimensional arrays and matrices, which are crucial when working with data obtained from physical measurements where the data structure is complex and requires fast and efficient traversal.

Django provides a powerful and robust framework for developing web applications. Built on top of Python, Django offers comprehensive tools for rapid and convenient development of web-based applications. An important aspect of this toolkit is the built-in data processing and the capability for storing and processing information in databases.

Combining Django with data processing tools like NumPy and pandas can be extremely useful for developing web applications that require data manipulation and analysis. Integrating these tools helps create powerful web-based applications that work with data from various sources and provide valuable analytical capabilities.

Additionally, Django provides capabilities for building public APIs, which is extremely useful for creating web applications that provide access to data through standardized interfaces. This facilitates sharing and integrating data between different systems and applications.

GraphQL is a powerful and flexible query language for working with data, used to interact with various data sources including databases, web services, and others. By employing this approach to data management, we achieve significant flexibility and control over data queries, allowing for precise extraction of only the data required for a specific task or application.

GraphQL not only simplifies data handling but also facilitates efficient interaction between the client and server sides of the application by providing optimized queries and responses. This abstract and flexible approach to data management, combined with data processing tools like NumPy and pandas, can be crucial in creating applications that require complex and dynamic data queries.

Using GraphQL has helped us build applications that not only provide user satisfaction but also facilitate data extraction and analysis, which is crucial in today's world of information and data.

REST (Representational State Transfer) is an architectural style for creating web services and APIs that provides a secure and efficient way to communicate between different systems and applications. REST APIs offer a lightweight and straightforward method for accessing resources through common HTTP requests, such as GET (for retrieving data), POST (for creating new data), PUT (for updating data), and DELETE (for deleting data).

REST APIs provide an easy way to access and manage data through standardized and intuitive requests, making it suitable for integration into various applications and platforms. This structured and transparent approach to data handling ensures flexibility and scalability of applications.

The cloud repository serves as a central storage for data coming from various sources. This includes information on pollutant concentrations, meteorological data, traffic, and many other factors affecting air quality. The cloud infrastructure provides secure storage and management of vast amounts of data.

Utilizing cloud infrastructure is crucial for processing large volumes of data and delivering services that require flexibility and scalability. Cloud platforms such as Amazon Web Services (AWS) and Microsoft Azure are used, providing resources and services for storing, processing, and analyzing data on a global scale. Combined with Python, Django, GraphQL, and REST APIs, along with cloud infrastructure, we provide tools for analyzing air quality and its impact on health.

4.5 Example of Using the Software Infrastructure: The AirLief Mobile Application

At the core of our software's functionality lies its ability to integrate data from GIS, measurements of air quality from stationary monitoring stations, and traffic data. This integration is achieved through advanced data merging techniques, allowing us to gain a comprehensive understanding of air quality on various spatial and temporal scales. GIS data provides essential geographic context, enabling us to precisely determine air quality measurements. By overlaying air quality and traffic data onto GIS layers, we can identify spatial patterns, hotspots, and correlations that would otherwise go unnoticed. This integration is crucial for our research as it allows us to assess the impact of traffic patterns and geographic characteristics on air quality levels.

The software has the capability to visualize air quality information on popular mapping platforms such as Google Maps and OpenStreetMap. This feature transforms raw data into a visually interpretable format, facilitating the dissemination of important information to a wide audience. We can analyze variations in air quality over time, identifying trends and anomalies to serve as the basis for our research findings. This temporal analysis provides insight into how air quality changes throughout different times of the day, seasons, or under certain meteorological conditions.

Integrating these diverse data sources into a cohesive framework enables us to draw significant conclusions about air quality and its interaction with the environment, contributing to the depth and comprehensiveness of our research.

AirLief is a mobile air quality application developed for iOS and Android platforms. It is a free, open-access application that shows users the air pollution levels around them based on their geolocation. AirLief retrieves data from the software system using the API created for data communication outside the system.

Air quality and atmospheric data are distributed from the software system to the mobile application through purpose-built infrastructure endpoints. To facilitate data transmission to our system, location information (GPS coordinates), geographic radius, or city inputs are entered, and real-time air quality data corresponding to the query is outputted. Three infrastructure endpoints have been created for the AirLief application:

– Nearest Point Measurements (input data: GPS coordinates): This endpoint allows users to receive results from the nearest monitoring station located near the specified GPS coordinates. It provides information about the current air quality at a specific location.

– Location-Based Measurements (input data: GPS coordinates, radius in km): This endpoint allows users to obtain the arithmetic mean values of measurements from all monitoring stations within a radius of specified kilometers around the given GPS coordinates. This gives a generalized overview of air quality in a particular area.

– City-Based Measurements (input data: city name): This endpoint enables users to obtain the arithmetic mean values of measurements from all monitoring stations in a specific city or populated place. It provides information about air quality within the boundaries of the specified city or populated place.

These infrastructure endpoints provide flexibility and convenience in extracting air pollution data, contributing to the extensive functionality of the air quality monitoring and management software system. They are presented to users in the mobile application in an easy and accessible manner.

The AirLief app retrieves and delivers air pollution information in an accessible way for over 100 countries from around 15,000 stations. The application itself is used in 70 countries by over 70,000 users, as shown in Figure 4.3. It consists of a dashboard that presents the current state of the air in an accessible and easy-to-understand manner; a map displaying values from stationary stations; the option to add favorite stations, and personalized tips for protection against air pollution. Figure 4.4 illustrates how the mobile application transforms raw data obtained from the software system into an Air Quality Index (AQI). The index, which directly correlates with the impact of air pollution on the human body, includes not only PM concentrations but also other pollutants such as ozone pollution, nitrogen dioxide, and others.

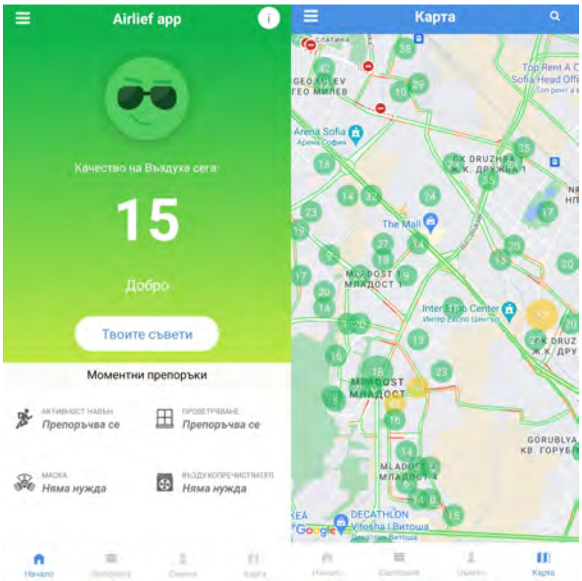


Figure 4.3: AirLief App is one of the applications developed from the software system, displaying real-time air pollution data from stationary sensors.

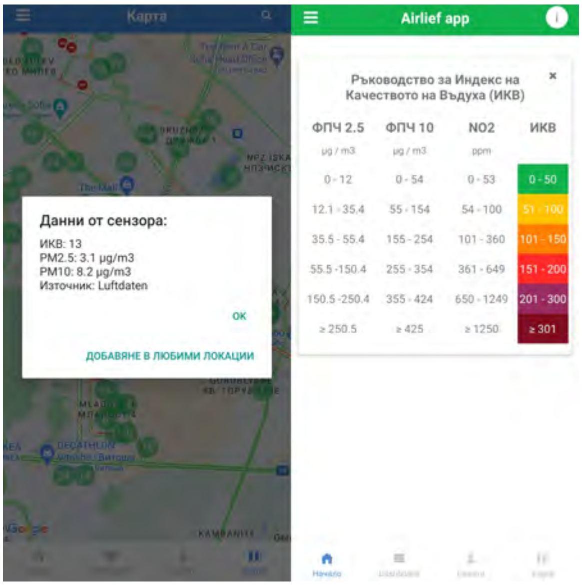


Figure 4.4: AirLief App - the mobile application that translates raw data into understandable Air Quality Index (AQI).

Chapter 5

Conclusion

The dissertation defines a goal and four tasks for achieving it. Initially, a detailed analysis was conducted on the data regarding air pollution and acute illnesses from two hospitals and emergency services in Sofia. The benefits of combining pollution data from official monitoring stations and laser sensor stations (low-cost IoT devices deployed by citizens) have been identified.

The dissertation develops and implements a solution for calibrating particulate matter (PM) data from laser stations using machine learning, which corrects errors and anomalies by comparing with data from a standardized station. Table **??** shows the improvement in laser station results, where the use of a combination of artificial neural networks and anomaly detection methods increases the R-squared value from 0.62 to 0.95 after calibration.

A route selection method based on PM pollution was developed, using a customized approach. Table 3.5 shows that in a field test with 10 cyclists, the routes generated by the software resulted in 67% better outcomes compared to other bicycle route-finding solutions.

All these tasks and solutions were developed into a scalable software system with a modular structure for collecting and processing data from various stations and IoT devices. This system facilitates effective integration with diverse hardware devices for measuring PM, utilizing different communication protocols. Practical results from the system's operation are visualized in a user interface, serving as a point of interaction between users and the system. The system is capable of processing and calibrating data as well as generating air pollution maps, and it is utilized in different projects, such as a mobile application accessed by thousands of users weekly.

The scientific and practical research presented in this dissertation provides the following contributions:

5.1 Scientific and Applied Contributions

1. A statistical correlation analysis was performed between air pollution and hospital and emergency admissions, identifying important cause-and-effect

relationships.

2. An algorithm for calibrating laser sensors was developed using a two-step method with artificial neural networks and anomaly detection. This innovative method supports more reliable and accurate measurements from laser sensors and has the potential to be used in a wide range of sensor applications.

3. An algorithm was created to evaluate the influence of humidity, altitude, and atmospheric pressure on air pollution data from laser sensors, allowing a more comprehensive and accurate understanding of the factors affecting air quality. It can also be applied in vertical planning.

4. An algorithm was developed to calculate the optimal cycling route based on PM concentration. This routing method promotes a healthy lifestyle and supports sustainable urban mobility.

5.2 Practical Contributions

1. Software for calibrating PM data from laser sensors was developed, using reference data from official monitoring stations. This software enables air quality control organizations to use more reliable data for decision-making and pollution management.

2. A software solution was developed to account for the influence of humidity, altitude, and atmospheric pressure, improving PM data from laser sensors. This software can be used by urban authorities and health organizations to monitor and manage air quality in real-time.

3. A software tool for finding the optimal cycling route based on PM inhalation was created. This software can be useful for individual cyclists and city infrastructures that promote cycling.

4. An IoT platform for aggregating and analyzing sensor data on air quality was developed. This platform supports automation and facilitates data integration from various sources, enhancing the efficiency and accuracy of air quality monitoring. In addition to meteorological data, it integrates traffic and GIS data. The platform visualizes maps, complemented by measurements from air quality sensors and real-time traffic data.

In light of the research and results presented in this dissertation, we can conclude that this project provides important scientific and practical contributions, influencing the development and improvement of air pollution measurement systems and related technologies.

5.3 Publications on the dissertation topic

- Zhivkov, P., & Simidchiev, A. (2024). Software Tool for Optimizing Cycling Route by Defining Cyclist Air Pollution Exposure
  Studies in Computational Intelligence, 1158 SCI, pp. 152–170, SJR 0.208 (2023)

  https://www.scopus.com/record/display.uri?eid=2-s2.0-85190651936&origin=resultslist&sort=plf-f&src=s&sid=3eca39362dd1584d77f2fb4c5901e3b2&sot=aut&sdt=a&sl=18&s=AU-ID%2857221769514%29&relpos=0&citeCnt=0&searchTerm=

- Zhivkov, P., & Simidchiev, A. (2022). Development of Software Tool for Optimization and Evaluation of Cycling Routes by Characterizing Cyclist Exposure to Air Pollution Annals of Computer Science and Information Systems, 32, 105-112.

  https://d1wqtxts1xzle7.cloudfront.net/99417790/230-libre.pdf?1677962458=&response-content-disposition=inline%3B+filename%3DDevelopment_of_Software_Tool_for_Optimiz.pdf&Expires=1725292377&Signature=QB4TSFkNK4~abulkZUncp_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

- Zhivkov, P. Optimization and Evaluation of Calibration for Low-cost Air Quality Sensors: Supervised and Unsupervised Machine Learning Models Annals of Computer Science and Information Systems, Volume 25, FedCSIS 2021, 2021, pp. 255–258

  https://www.scopus.com/inward/record.uri?eid=2-s2.0-85117793346&doi=10.15439%2f2021F95&partnerID=40&md5=2f0e02da6b0d78aaa6523df4f48308c9

- Zhivkov, P., & Simidchiev, A. (2020). Relationship between Particulate Matter and Health Indicators for Acute Morbidity in Sofia.
  In Proceeding of 1st international Conference on Environmental Protection and Disaster Risks (pp. 180-193).

  https://www.ceeol.com/search/chapter-detail?id=926651

5.4 Declaration of originality

I declare that the dissertation contains original results, obtained during research conducted by me with the support and assistance of my supervisor.

The results that have been obtained, described and/or published by other scientists are correctly and extensively cited in the bibliography.

This dissertation has not been applied for a degree at another college, university or research institute.

Signature:

Bibliography

[1] Norfazillah Ab Manan, Azimatun Noor Aizuddin, and Rozita Hod. Effect of air pollution and hospital admission: a systematic review. *Annals of global health*, 84(4):670, 2018.

[2] Stacey E Alexeeff, Noelle S Liao, Xi Liu, Stephen K Van Den Eeden, and Stephen Sidney. Long-term pm2. 5 exposure and risks of ischemic heart disease and stroke events: review and meta-analysis. *Journal of the American Heart Association*, 10(1):e016890, 2021.

[3] Emma Barnes and Marc Schlossberg. Improving cyclist and pedestrian environment while maintaining vehicle throughput: Before-and after-construction analysis. *Transportation research record*, 2393(1):85–94, 2013.

[4] Hannah Bast, Daniel Delling, Andrew Goldberg, Matthias Müller-Hannemann, Thomas Pajor, Peter Sanders, Dorothea Wagner, and Renato F Werneck. Route planning in transportation networks. In *Algorithm engineering*, pages 19–80. Springer, 2016.

[5] M.L. Bell, J.M. Samet, and F. Dominici. Time-series studies of particulate matter. *Annu. Rev. Public Health*, 25:247–280, 2004.

[6] Alexander Y Bigazzi and Miguel A Figliozzi. Review of urban bicyclists' intake and uptake of traffic-related air pollution. *Transport Reviews*, 34(2):221–245, 2014.

[7] Iain M Carey, Richard W Atkinson, Andrew J Kent, Tjeerd Van Staa, Derek G Cook, and H Ross Anderson. Mortality associations with long-term exposure to outdoor air pollution in a national english cohort. *American journal of respiratory and critical care medicine*, 187(11):1226–1233, 2013.

[8] N. Castell et al. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment international*, 99:293–302, 2017.

[9] J.W. Cherrie et al. Effectiveness of face masks used to protect beijing residents against particulate air pollution. *Occupational and environmental medicine*, 75(6):446–452, 2018.

[10] Theodoros Chondrogiannis, Panagiotis Bouros, Johann Gamper, and Ulf Leser. Alternative routing: k-shortest paths with limited overlap. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–4, 2015.

[11] Theodoros Chondrogiannis, Panagiotis Bouros, Johann Gamper, and Ulf Leser. Exact and approximate algorithms for finding k-shortest paths with limited overlap. In *20th International Conference on Extending Database Technology: EDBT 2017*, pages 414–425, 2017.

[12] L. Clancy et al. Effect of air-pollution control on death rates in dublin, ireland: an intervention study. *The lancet*, 360(9341):1210–1214, 2002.

[13] Aaron J Cohen, Michael Brauer, Richard Burnett, H Ross Anderson, Joseph Frostad, Kara Estep, Kalpana Balakrishnan, Bert Brunekreef, Lalit Dandona, Rakhi Dandona, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *The Lancet*, 389(10082):1907–1918, 2017.

[14] Audrey De Nazelle, Mark J Nieuwenhuijsen, Josep M Antó, Michael Brauer, David Briggs, Charlotte Braun-Fahrlander, Nick Cavill, Ashley R Cooper, Hélène Desqueyroux, Scott Fruin, et al. Improving health through policies that promote active travel: a review of evidence to support integrated health impact assessment. *Environment international*, 37(4):766–777, 2011.

[15] CE Delft. Health costs of air pollution in european cities and the linkage with transport. *CE Delft: Delft, The Netherlands*, 2020.

[16] Inês D Do Vale, Ana S Vasconcelos, and Gonçalo O Duarte. Inhalation of particulate matter in three different routes for the same od pair: A case study with pedestrians in the city of lisbon. *Journal of Transport & Health*, 2(4):474–482, 2015.

[17] F. Dominici, L. Sheppard, and M. Clyde. Health effects of air pollution: a statistical review. *International Statistical Review*, 71(2):243–276, 2003.

[18] Jiyuan Dong, Yurong Liu, and Hairong Bao. Revalue associations of short-term exposure to air pollution with respiratory hospital admissions in lanzhou, china after the control and treatment of current pollution. *International Journal of Hygiene and Environmental Health*, 231:113658, 2021.

[19] M. Gochfeld and J. Burger. Disproportionate exposures in environmental justice and other populations: the importance of outliers. *American Journal of Public Health*, 101(S1):S53–S63, 2011.

[20] T. Guan et al. The effects of facemasks on airway inflammation and endothelial dysfunction in healthy young adults: a double-blind, randomized, controlled crossover study. *Particle and fibre toxicology*, 15(1):30, 2018.

[21] A. Gábelová et al. Genotoxicity of environmental air pollution in three european cities: Prague, košice and sofia. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 563(1):49–59, 2004.

[22] Walter Ham, Abhilash Vijayan, Nico Schulte, and Jorn D Herner. Commuter exposure to pm2. 5, bc, and ufp in six common transport microenvironments in sacramento, california. *Atmospheric Environment*, 167:335–345, 2017.

[23] Ghassan B Hamra, Neela Guha, Aaron Cohen, Francine Laden, Ole Raaschou-Nielsen, Jonathan M Samet, Paolo Vineis, Francesco Forastiere, Paulo Saldiva, Takashi Yorifuji, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environmental health perspectives*, 2014.

[24] A.J. Hedley et al. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in hong kong: an intervention study. *The lancet*, 360(9346):1646–1652, 2002.

[25] Mario A Hernández, Omar Ramírez, John A Benavides, and Juan F Franco. Urban cycling and air quality: Characterizing cyclist exposure to particulate-related pollution. *Urban Climate*, 36:100767, 2021.

[26] Jalil Jaafari, Kazem Naddafi, Masud Yunesian, Ramin Nabizadeh, Mohammad Sadegh Hassanvand, Mansour Shamsipour, Mohammad Ghanbari Ghozikali, Hamid Reza Shamsollahi, Shahrokh Nazmara, and Kamyar Yaghmaeian. The acute effects of short term exposure to particulate matter from natural and anthropogenic sources on inflammation and coagulation markers in healthy young adults. *Science of The Total Environment*, 735:139417, 2020.

[27] Rohan Jayaratne, Xiaoting Liu, Phong Thai, Matthew Dunbabin, and Lidia Morawska. The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. *Atmospheric Measurement Techniques*, 11(8):4883–4890, 2018.

[28] Marilena Kampa and Elias Castanas. Human health effects of air pollution. *Environmental pollution*, 151(2):362–367, 2008.

[29] K. Katsouyanni et al. Short term effects of air pollution on health: a european approach using epidemiologic time series data: the aphea protocol. *Journal of Epidemiology & Community Health*, 50(Suppl 1):S12–S18, 1996.

[30] Christine M Kendrick, Adam Moore, Ashley Haire, Alexander Bigazzi, Miguel Figliozzi, Christopher M Monsere, and Linda George. Impact of bicycle lane characteristics on exposure of bicyclists to traffic-related particulate matter. *Transportation research record*, 2247(1):24–32, 2011.

[31] P. Li et al. The acute effects of fine particles on respiratory mortality and morbidity in beijing, 2004–2009. *Environmental Science and Pollution Research*, 20(9):6433–6444, 2013.

[32] Feng Lu, Dongqun Xu, Yibin Cheng, Shaoxia Dong, Chao Guo, Xue Jiang, and Xiaoying Zheng. Systematic review and meta-analysis of the adverse health effects of ambient pm2. 5 and pm10 pollution in the chinese population. *Environmental research*, 136:196–204, 2015.

[33] Lisha Luo, Yunquan Zhang, Junfeng Jiang, Hanghang Luan, Chuanhua Yu, Peihong Nan, Bin Luo, and Mao You. Short-term effects of ambient air pollution on hospitalization for respiratory disease in taiyuan, china: a time-series analysis. *International journal of environmental research and public health*, 15(10):2160, 2018.

[34] Kamal Jyoti Maji, Wei-Feng Ye, Mohit Arora, and SM Shiva Nagendra. Pm2. 5-related health and economic loss assessment for 338 chinese cities. *Environment international*, 121:392–403, 2018.

[35] Ernesto QV Martins and Marta Pascoal. A new implementation of yen's ranking loopless paths algorithm. *Quarterly Journal of the Belgian, French and Italian Operations Research Societies*, 1(2):121–133, 2003.

[36] L. Morawska et al. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environment international*, 116:286–299, 2018.

[37] Christopher JL Murray. The global burden of disease study at 30 years. *Nature medicine*, 28(10):2019–2026, 2022.

[38] National Statistical Institute of Bulgaria (NSI). Population by cities https://www.nsi.bg/en/content/2981/population-towns-and-sex, 2019.

[39] World Health Organization. *International classification of diseases:[9th] ninth revision, basic tabulation list with alphabetic index.* 1978.

[40] World Health Organization. International classification of diseases (icd) information sheet. 2018.

[41] Luc Int Panis, Bas de Geus, Grégory Vandenbulcke, Hanny Willems, Bart Degraeuwe, Nico Bleux, Vinit Mishra, Isabelle Thomas, and Romain Meeusen. Exposure to particulate matter in traffic: a comparison of cyclists and car passengers. *Atmospheric Environment*, 44(19):2263–2270, 2010.

[42] Hye-Youn Park, Susan Gilbreath, and Edward Barakatt. Respiratory outcomes of ultrafine particulate matter (ufpm) as a surrogate measure of near-roadway exposures among bicyclists. *Environmental Health*, 16(1):1–7, 2017.

[43] S. Pattenden, B. Nikiforov, and B. Armstrong. Mortality and temperature in sofia and london. *Journal of Epidemiology & Community Health*, 57(8):628–633, 2003.

[44] C.A. Pope, D.W. Dockery, and J. Schwartz. Review of epidemiological evidence of health effects of particulate air pollution. *Inhalation toxicology*, 7(1):1–18, 1995.

[45] Sanjay Rajagopalan, Sadeer G Al-Kindi, and Robert D Brook. Air pollution and cardiovascular disease: Jacc state-of-the-art review. *Journal of the American College of Cardiology*, 72(17):2054–2070, 2018.

[46] A.M. Rendón et al. Effects of urbanization on the temperature inversion breakup in a mountain valley with implications for air quality. *Journal of Applied Meteorology and Climatology*, 53(4):840–858, 2014.

[47] J. Samet and D. Krewski. Health effects associated with exposure to ambient air pollution. *Journal of toxicology and environmental health, Part A*, 70:227–242, 2007.

[48] J.M. Samet et al. The national morbidity, mortality, and air pollution study. part ii: morbidity and mortality from air pollution in the united states. *Res Rep Health Eff Inst*, 94(pt 2):5–79, 2000.

[49] J. Shi et al. Cardiovascular benefits of wearing particulate-filtering respirators: a randomized crossover trial. *Environmental health perspectives*, 125(2):175–180, 2017.

[50] S. Shin et al. Ambient air pollution and the risk of atrial fibrillation and stroke: A population-based cohort study. *Environ Health Perspect*, 127(8):87009, 2019.

[51] L.A. Tallon et al. Cognitive impacts of ambient air pollution in the national social health and aging project (nshap) cohort. *Environ Int*, 104:102–109, 2017.

[52] Yan Tao, Shengquan Mi, Shuhong Zhou, Shigong Wang, and Xiaoyun Xie. Air pollution and hospital admissions for respiratory diseases in lanzhou, china. *Environmental pollution*, 185:196–201, 2014.

[53] G.D. Thurston et al. A joint ers/ats policy statement: what constitutes an adverse health effect of air pollution? an analytical framework. *European Respiratory Journal*, 49(1), 2017.

[54] H. Traboulsi et al. Inhaled pollutants: the molecular scene behind respiratory and systemic diseases associated with ultrafine particulate matter. *International Journal of Molecular Sciences*, 18(2):243, 2017.

[55] J. Vermylen et al. Ambient air pollution and acute myocardial infarction. *Journal of Thrombosis and Haemostasis*, 3(9):1955–1961, 2005.

[56] W. Wang et al. Particulate air pollution and ischemic stroke hospitalization: How the associations vary by constituents in shanghai, china. *Sci Total Environ*, 695:133780, 2019.

[57] Scott Weichenthal, Ryan Kulka, Aimee Dubeau, Christina Martin, Daniel Wang, and Robert Dales. Traffic-related air pollution and acute changes in heart rate variability and respiratory function in urban cyclists. *Environmental health perspectives*, 119(10):1373–1378, 2011.

[58] Lingkun Wu, Xiaokui Xiao, Dingxiong Deng, Gao Cong, Andy Diwen Zhu, and Shuigeng Zhou. Shortest path and distance queries on road networks: An experimental evaluation. *arXiv preprint arXiv:1201.6564*, 2012.

[59] Sherrie Xie and Blanca E Himes. Personal environmental monitoring. In *Precision in Pulmonary, Critical Care, and Sleep Medicine*, pages 305–320. Springer, 2020.

[60] Antonella Zanobetti, Francesca Dominici, Yun Wang, and Joel D Schwartz. A national case-crossover analysis of the short-term effect of pm2. 5 on hospitalizations and mortality in subjects with diabetes and neurological disorders. *Environmental Health*, 13(1):1–11, 2014.

[61] R. Zhang et al. Acute effects of particulate air pollution on ischemic stroke and hemorrhagic stroke mortality. *Front Neurol*, 9:827, 2018.

[62] Petar Zhivkov. Optimization and evaluation of calibration for low-cost air quality sensors: Supervised and unsupervised machine learning models. In *2021 16th Conference on Computer Science and Intelligence Systems (FedCSIS)*, pages 255–258. IEEE, 2021.

[63] Petar Zhivkov and Alexander Simidchiev. Development of software tool for optimization and evaluation of cycling routes by characterizing cyclist exposure to air pollution. *Annals of Computer Science and Information Systems*, 32:105–112, 2022.

[64] Petar Zhivkov and Alexander Simidchiev. Software tool for optimizing cycling route by defining cyclist air pollution exposure. In *The Workshop on Computational Optimization*, pages 152–170. Springer, 2024.

[65] N. Zimmerman et al. Closing the gap on lower cost air quality monitoring: Machine learning calibration models to improve low-cost sensor performance. *Atmos. Meas. Tech. Discuss*, pages 1–36, 2017.

[66] Moniek Zuurbier, Gerard Hoek, Marieke Oldenwening, Virissa Lenters, Kees Meliefste, Peter Van Den Hazel, and Bert Brunekreef. Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route. *Environmental health perspectives*, 118(6):783–789, 2010.

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

АВТОРЕФЕРАТ НА ДИСЕРТАЦИЯ

за присъждане на образователна и научна степен "доктор" по докторска програма „Информатика"

МОДЕЛИРАНЕ НА КАЧЕСТВОТО НА АТМОСФЕРНИЯ ВЪЗДУХ ВЪРХУ ЗДРАВНИ И ИКОНОМИЧЕСКИ АСПЕКТИ

Петър Руменов Живков

Ръководител: Проф. Стефка Фиданова

Научно жури:

Проф. Иван Димов
Проф. Олимпия Роева
Доц. Райна Георгиева
Проф. Златинка Ковачева
Проф. Евдокия Сотирова

Институт по информационни и комуникационни технологии

Секция „Паралелни алгоритми и машинно обучение с Лаборатория по невротехнологии"

Глава 1

Въведение

Замърсяването на въздуха и по-конкретно фините прахови частици (ФПЧ), е водещата екологична причина за заболеваемост и преждевременна смъртност, след като през 1990 г.е било на пета позиция [48]. Смъртните случаи, причинени от ФПЧ в околната среда, са нараснали през последните 25 години [15], а излагането на по-високо замърсяване на атмосферния въздух е свързано с по-висока смъртност [9]. А последният доклад за оцен

ФПЧ се класифицират въз основа на техния размер, като частиците с размери по-малки от 10 микрометра (ФПЧ10) и 2,5 микрометра (ФПЧ2,5) предизвикват най-голямо безпокойство поради способността им да проникват дълбоко в белите дробове и да причиняват неблагоприятни последици за здравето. ФПЧ представляват сложна смес от твърди частици и течни капки, суспендирани във въздуха, с различн размери и химичен състав. Тези частици могат да произхождат от природни източници като прах и полени, както и от човешки дейности като промишлени емисии, отработени газове от превозни средства и изгаряне на биомаса и твърди горива.

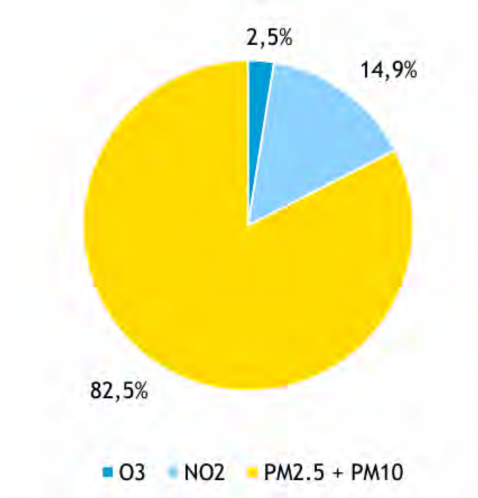
Според проучвания излагането на ФПЧ2,5 и ФПЧ10 е свързано с ранна смъртност от няколко заболявания, като сърдечносъдови, респираторни, рак на белия дроб и инфекции на горните и долните дихателни пътища [2], [28],[32]. В това число замърсяването на атмосферния въздух в големите градове увеличава риска от белодробен и системен оксидативен стрес, хипоксемия,имунологични модификации, атеросклероза и ускорена прогресия на сърдечносъдовите заболявания и хроничната обструктивна белодробна болест (ХОББ) [43], а епидемиологични проучвания показват, че излагането на ФПЧ увеличава риска от диабет [74].

Тежестта на последиците за здравето може да варира в зависимост от концентрацията и продължителността на излагане, както и от здравословното състояние на индивида и неговата чувствителност към замърсителите. Някои групи от населението, като децата, възрастните хора и тези с вече съществуващи респираторни или сърдечносъдови заболявания, са особено уязвими към здравните ефекти от замърсяването на въздуха [58]

Изследвания показват силна връзка между замърсяването на въздуха и хоспитализациите за респираторни заболявания [66, 21, 44]. А кохортен преглед на 175 статии обобщава, че излагането на мръсен въздух увеличава вероятността от развитие на различни заболявания като респираторни заболявания (астма, хронична обструктивна белодробна болест (ХОББ) и рак на белия дроб), съдови заболявания (инфаркти, инсулти и хипертония), бъбречни, чернодробни и други [1].

1.1 Актуалност и мотивировка на темата

Нарастващата загриженост по отношение на въздействието на замърсяването на въздуха върху общественото здраве налага разработването на прогнозни модели за предвиждане нивото на остра заболеваемост въз основа на данните за качеството на въздуха [45]. В последния доклад на Европейския Съюз (ЕС) за качеството на въздуха се оценява, че усилията на ЕС за подобряване на качеството му са само "частично ефективни"и специално изтъква ,че действията за намаляване на ФПЧ2,5 са недостатъчни. На фигура 1.1 е представен резултат от проучване с 432 града в ЕС, което анализира икономическите последствия от трите основни замърсители в европейските градове: ФПЧ, азотен диоксид и озон [18]. Резултатите показват, че ФПЧ са основната причина (82,5 %) за икономическите щети в следствие на болести и смъртност от мръсен въздух. Изследването използва програмата за моделиране AIRQ+, създадена от СЗО, като включва стойности за замърсяване и икономически данни за всеки от градовете.



Фигура 1.1: Принос на трите основни градски замърсители към общите щети за заболеваемост и смъртност в 432 европейски града

В светлината на тези констатации, нараства необходимостта от модели, които да могат да изследват заболяваемостта при остри състояния, базирани на данни за качеството на въздуха. Въпреки нарастващата осведоменост за рисковете за здравето, свързани със замърсяването на въздуха, съществува необходимост от задълбочен анализ на конкретните замърсители и тяхното въздействие върху различни здравни резултати. Съществуващите проучвания показват връзки между замърсяването на въздуха и респираторните заболявания, сърдечносъдовите заболявания, неблагоприятните последици при раждане и дори проблемите с психичното здраве. Въпреки това все още е необходимо цялостно проучване на мащаба и механизмите на тези връзки, за да се информират основаните на доказателства интервенции и политики на държавно и локално ниво.

У нас нивата на ФПЧ често надхвърлят нормите на ЕС за качество на въздуха. ЕК през април 2017 г. спечели дело срещу България заради наднормено замърсяване на въздуха в редица градове. Също така България е лидер по загубени години на живот в добро здраве, заради мръсен въздух. Фигура според международен доклад, гражданите у нас скъсяват с 2,5 години живота си в следствие на възушно замърсяване, докато средно за Европейския съюз (ЕС) този показател е 0,7 години (`https://www.eca.europa.eu/Lists/ECADocuments/SR18_23/SR_AIR_QUALITY_BG.pdf`). Качеството на въздуха превишава нормите основно поради битовото отопление, автомобилния трафик и географските фактори, които задържат замърсителите в атмосферата. Липсата на солидни изследвания и статистически данни за здравните ефекти от замърсяването с ФПЧ в София подкопава способността на града да разработва основани на доказателства стратегии за намаляване на въздействието му. Без точна и подробна информация създателите на политики и служителите в областта на общественото здраве са изправени пред значителни предизвикателства при прилагането на целенасочени интервенции и повишаването на обществената осведоменост относно сериозността на проблема. Ето защо е налице спешна необходимост от всеобхватни изследователски усилия и подобрени методи за събиране на данни, които да преодолеят съществуващите пропуски в знанията и да проправят пътя за вземане на информирани решения в борбата със замърсяването с ФПЧ и за подобряване на качеството на здравето в София.

1.2 Цел и задачи на дисертационния труд

Целта на дисертационния труд е изследване влиянието на фините прахови частици (ФПЧ) върху острите заболявания в София и намиране на методи за превенция.

За постигането на тази цел са формулирани следните 4 задачи:

- Връзка между фините прахови частици и здравните показатели за остра заболеваемост в София;
- Подобряване на данни от граждански станции за измерване качеството на въздуха чрез калибриране с машинно обучение: двустъпков метод;
- Разработване на софтуерен инструмент за оптимизиране и оценка на велосипедни маршрути чрез характеризиране на излагането на велосипедистите на замърсяване на въздуха;
- Разработване на IoT платформа за агрегация и моделиране на сензорни данни за качество на въздуха.

1.3 Методология на изследването

В настоящия дисертационен труд е използван мултидисциплинарен подход, съчетаващ литературен обзор, количествен анализ и усъвършенствано статистическо моделиране. Извършен е систематичен преглед на съществуващи изследвания, за да се обобщят и синтезират настоящите доказателства за връзката между качеството на въздуха и здравните резултати. Освен това са анализирани епидемиологични проучвания и данни от агенции за здравно наблюдение, за да се определи количествено връзката между конкретни замърсители и последиците за здравето. Използвани са съвременни статистически техники, като регресионно моделиране и пространствен анализ, за да се изследват сложните взаимодействия и пространствените модели на замърсяването на въздуха и здравните последици.

Методологията, използвана в този дисертационен труд, включва систематичен подход за изследване на въздействието на фините прахови частици върху острите заболявания в София и разработване на стратегии за смекчаване на последиците. Методологията включва събиране на данни от болници и измервателни станции, анализ на данните с цел установяване на корелации, калибриране на граждански лазерни станции, разработване на IoT платформа за визуализация на данните и оптимизиране на велосипедни алеи.

Събирането на данни за постъпленията и хоспитализациите от две големи болници в София, данни от регистъра на бърза помощ и измервания на ФПЧ както от официални измервателни станции на Изпълнителната агенция по околната среда (ИАОС), така и от граждански лазерни станции. Събрани са и метеорологични данни, включително влажност, атмосферно налягане и температура, данни за трафик и други.

Приложени са техники за анализ на данните, за да се изследва връзката между нивата на ФПЧ и острите заболявания. Извършен е статистически анализ и сравнения на нивата на ФПЧ под и над здравните стандарти на Световната здравна организация (СЗО). Процесът на калибриране се съсредоточи върху калибрирането на данните от гражданските лазерни станции, като са използвани модели за машинно обучение с контролиран и безконтролен метод. За точното калибриране са използвани референтни данни от измервателни станции на МОСВ, като са взети в предвид влиянието на фактори като влажност, атмосферно налягане и температура. Разработен е модел за оптимизиране на велосипедните трасета за минимално вдишване на ФПЧ. Това включва използване на модифициран алгоритъм за намиране на най-краткия път и провеждане на реални полеви тестове за валидиране на методологията. Разработена е софтуерна система, която визуализира местоположението и показателите на измервателните станции върху географска карта. Приложен е нов подход за агрегиране, организиране, обработка, моделиране и обмен на данни в IoT системата.

В процеса на изследване на проблема е използван широк сфера от литература. Създадените нови методи и алгоритми са публикувани в [[**?**], [76], [77], [78]]. Представената методология позволява цялостен анализ на въздействието на ФПЧ върху острите заболявания, калибриране на гражданските лазерни станции чрез машинно обучение и невронни мрежи и оптимизиране на велосипедни маршрути с оглед на замърсяването от ФПЧ. В съответствие с поставените задачи са разработени алгоритми и методологии решаващи конкретните проблеми. Направена е софтуерно приложение на всеки от разработените алгоритм и визуализиране на данните чрез IoT платформа. Платформата и софтурът са написани на езика Python и използвайки рамката Django.

Глава 2

Връзка между замърсяването с ФПЧ и здравни индикатори за остра заболеваемост

Връзките между градското въздушно замърсяване (ВЗ) и човешкото здраве са последователно и ясно установени от редица изследователи [67], [60], [41], като открояващи групи забоялвания са сърдечно-съдови [69], невроваскуларни [65], [11] и белодробни [68]. Според скорошно усъвършенствано моделиране се предполага, че има близо 9 милиона смъртни случая годишно от ВЗ [55]. Около 25% от преждевременните смъртни случаи, свързани с ВЗ, са респираторни по презумпция [6].

Значителна литература за епидемиологични проучвания предполага връзка между острата заболеваемост и излагането на замърсяване на въздуха от ФПЧ [20]. Повечето от тези данни идват от анализи на времеви серии [24], сравняващи вариациите в хоспитализацията със средните вариации на прахови частици [29]. ВЗ е отговорно за възпаленията на дихателните пътища [14] и като се има предвид, че дихателната система е обща врата за влизане, минимизирането на експозицията на дихателните пътища също минимизира сърдечно-съдовите проблеми [63]. Освен това съществуват проучвания, които моделират замърсяването на множество градове, като Европейското замърсяване на въздуха и здравето: европейски подход (APHEA) [36] и Американското национално изследване на заболеваемостта, смъртността и замърсяването на въздуха (NMMAPS) [61], като и двете предоставят последователни доказателства за връзка здраве и замърсители на въздуха за множество градове чрез покриване на голяма географска област.

Последиците от замърсяването на въздуха могат да се разглеждат като повишаване на риска от заболяване или нараняване на дадено лице или като допълнителен общ риск за благосъстоянието на населението [23]. Целта на управлението на качеството на въздуха е да контролира или избягва неблагоприятните въздействия на замърсяването на въздуха върху общественото здраве. Следователно е важно, чрез стандартите за ФПЧ, да се дефинират такива въздействия, които се считат за „неблагоприятни“, и да се разграничат от тези въздействия, които не се считат за неблагоприятни, като по този начин усилията за защита се концентрират върху замърсителите, които причиняват най-екстремни въздействия върху здравето.

Таблица 2.1: Замърсяване на въздуха: препоръки на СЗО и законодателство на ЕС (в $\mu g/m^3$)

Норми за замърсяване на въздуха	СЗО		ЕС	
	ФПЧ10	ФПЧ2,5	ФПЧ10	ФПЧ2,5
среднодневна	50	25	50	–
средногодишна	20	10	40	25

2.1 Стандарти за ФПЧ: Анализ на законодателството на ЕС и насоки на СЗО

Законодателната рамка на ЕС за нивата на ФПЧ във въздуха се основава на Директива за Качество на Въздуха (2008/50/ЕО) и се допълва от Директива за Национални Емисионни Прагове (2016/2284/ЕС). Тези регулации задължават държавите-членки да изпълняват стандартите за ФПЧ във въздуха в съответствие със специфични прагови стойности, извлечени от научни оценки. Съответствието се наблюдава от Европейската комисия, а неспазването може да доведе до правни действия и глоби. В сравнение, СЗО предоставя общи указания за качество на въздуха без законодателни задължения, оставяйки изпълнението на индивидуалните държави.

В сравнителния анализ между законодателството на ЕС и Указанията на СЗО относно нивата на ФПЧ във въздуха се изявяват няколко ключови разлики (Таблица 2.1). Докато ЕС и СЗО и двете препоръчват дневен среден концентрационен праг за ФПЧ10 от 50 $\mu g/m^3$, ЕС и СЗО се различават значително в годишните си средни препоръки, като стандартите на СЗО са двойно по-строги строги (20 $\mu g/m^3$, срещу 40 $\mu g/m^3$ от ЕС). За ФПЧ2,5, ЕС нямат концентрационен праг за средендневни норми и има значителни разлики в годишните препоръки, като СЗО настоява за по-строги лимити. Освен това стандартите на ЕС са специфични за региона, докато указанията на СЗО са необвързващи и универсални, предоставяйки насоки на страни по целия свят.

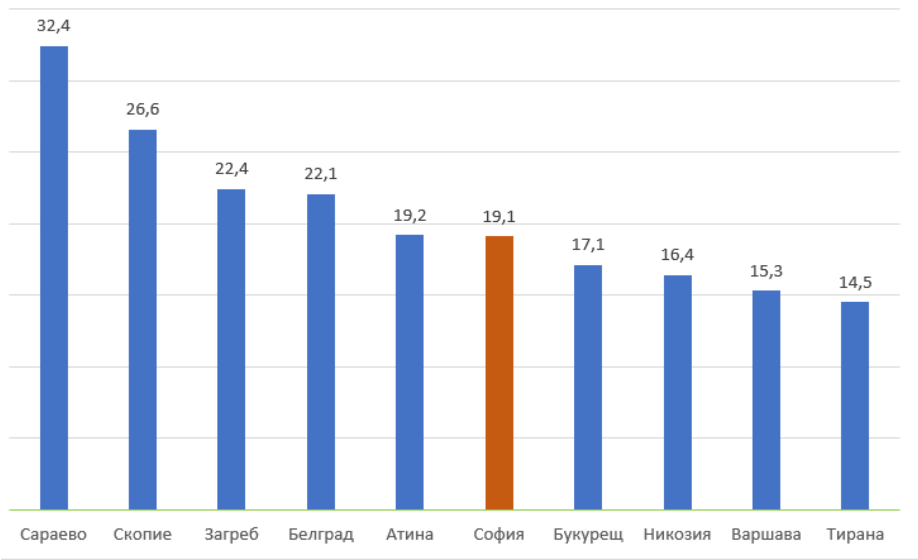
За целите на тази дисертация ще използваме насоките за качество на въздуха на СЗО, тъй като те са препоръчителни в световен мащаб, имат ограничение за дневни норми на ФПЧ2,5 и се базират на чисто здравни аспекти, в сравнение с директивата на ЕС, която е с фокус към регионални и национални нужди и включва политико-икономически и социални аспекти в директивите.

В последния си доклад СЗО полага специално внимание на замърсяването от ФПЧ и смята излагането на ФПЧ 2,5 за най-големия убиец свързан с въздуха в световен план. Около 80 % от смъртните случаи в световен мащаб, дължащи се на излагане на ФПЧ2,5, могат да бъдат предотвратени, ако страните спазват годишна прагова стойност за ФПЧ2,5 от 5 $\mu g/m^3$. Постигането на междинните цели също ще има значителни ползи за здравето. В случая с ФПЧ2,5 постигането на междинна цел 4 (10 $\mu g/m^3$) би довело до приблизително 48% намаление на общия брой смъртни случаи, дължащи се на експозиция на ФПЧ2,5.

2.2 Географска зона на изследване

София е разположена в котловина и се характеризира с високо количество антропогенни емисии и честа поява на застойни метеорологични условия. Градът е с население от 1,2 милиона души [49] и е разположен в Софийското котловинно поле. Районът е признат за проблемно място, където особено през зимата има многобройни превишения на нормите в европейското законодателство за контрол на замърсяването на въздуха.

От години София се бори със значителни нива на замърсяване на въздуха, като ФПЧ постоянно надвишават препоръчителните граници, определени от СЗО и ЕС. Тези повишени нива на ФПЧ във въздуха на града представляват сериозна заплаха за здравето и благосъстоянието на неговите жители. На Фигура 2.2 е показана средна концентрация на ФПЧ2,5 в най-замърсените столици в Европа през 2022 г. На нея се вижда, че София заема една от призовите позиции в тази класация. Също важен е фактът, че данните за измерване на качеството на въздуха са недостатъчни за София. За нивата на ФПЧ 2.5 в столицата се подават данни от една единствена станция - АИС Хиподрума, което е недостатъчно за качествени статистически анализи за град с размерите и населението на София.



Фигура 2.1: Средна концентрация на ФПЧ2,5 в най-замърсените столици в Европа през 2022 г. (в микрограми на кубичен метър въздух)

Следователно тази разлика между качеството на въздуха в летни и зимни условия ще бъде добър модел за тестване на нашата хипотеза за значителна връзка между замърсяването и последиците за здравето в София. Литературата изобилства с модели, които могат да се използват за оценка на хипотезата за замърсяване/здраве. Конкретно за София не са правени подобни проучвания. Корелация между температурата и смъртността около София се прави с помощта на линейни и нелинейни модели [54]. Други изследвания, които изследват генотоксичността на атмосферния въздух в 3 европейски града, включително София, показват, че замърсяването на въздуха през зимата е от 6 до 10 пъти по-високо в сравнение с летния въздух [25].

2.3 Методология на изследването

До колкото е известно това е първото подобно изследване с реални данни от официални и граждански източници на информация за замърсяването на въздуха с прахови частици, изчерпателни данни от дейността на центъра за спешна медицинска помощ (бърза помощ), данни от две от най-големите болници, свързани с достъпа до спешна помощ в София -УМБАЛСМ Пирогов и УМБАЛ Токуда. Резултатите могат да бъдат сравнени и контрастирани с други международни проучвания с местни данни.

Целта на това изследване е да се установи коя остра заболеваемост и с колко се увеличава в дните, когато нивата на ФПЧ не отговарят на насоките на СЗО, като се използва анализ на времеви серии.

Решени са следните четири подзадачи: сравняване на данните за качеството на въздуха със здравните данни, получени от болници и бърза помощ; сравнителен анализ и подчертаване на ключовите проблемни области; тестване на хипотезата, че данните от евтини сензори могат да бъдат полезни при такива епидемиологични изследвания; идентифициране на бъдещи перспективи и обобщение на ключовите области, в които са необходими допълнителни изследвания за подобряване на ефективността на модела.

Данни за замърсяването на въздуха

Първо ще открием кои дни, в рамките на наблюдавания период, концентрациите на ФПЧ надвишават средно дневните лимити на СЗО. След това ще сравним здравните данни за дните, в които качеството на въздуха е в рамките на указанията на СЗО, с дните, които не са.

Данните за качеството на въздуха под формата на средночасови концентрации на ФПЧ2,5 и ФПЧ10 са за 2018 г. и 2019 г. и представляват независими серии за всеки елемент от официални сензори, предоставени от (а) Изпълнителната агенция по околна среда (ИАОС) и (б) лазерни станции от мрежата luftdaten (глобална сензорна мрежа, управлявана от доброволци, която създава отворени данни за околната среда).

Данните се събират от 5 официални наблюдателни площадки (Дружба, Надежда, Хиподрума, Павлово и Младост). За измерване на ФПЧ10 и ФПЧ2,5 се използва метод за бета атенюация съгласно Европейската директива 2008/50/ЕС (Директива, 2008). Според предишни проучвания използването на данни за качеството на въздуха от евтини лазерни станции има за цел да увеличи потенциалните ползи от традиционните мрежи за наблюдение с допълнителна географска и времева резолюция на измерване [10]. Освен това евтините станции могат да надградят официалните измервателни уреди и по-специално ограничено количество данни и разпространение на станции [79]. Смята се, че евтините сензори са подходящи за много специфични цели, включително разширяване на разговорите с общностите и общата гражданска осведоментост по проблема на ВЗ [47]. В София към момента на изследването има няколко граждански мрежи от лазерни станции, които надграждат, с допълнителна географска и времева разделителна способност на измерване, официалните станции на ИАОС.

Здравни данни

Поради доброволния характер на предоставянето на данни има различни времеви характеристики и формати от различни източници:

- Обобщени данни за дейността на Центъра за спешна медицинска помощ София по диагнози на центъра - от 01.01.2017 г. до 14.03.2019 г.;
- Информация за дейността по диагностики от ДКЦ Токуда - 02.01.2018 г. до 31.12.2018 г.;
- Информация за дейността по диагностика от УМБАЛСМ Пирогов - 01.01.2018 г. до 31.05.2019 г.;
- Информация за дейността на хоспитализираните диагностики от Пирогов - 01.01.2018 г. до 31.12.2018 г.

Изследването използва Международната класификация на болестите (ICD) и по-конкретно ICD-10 за сегментиране на заболяванията и идентифицираната заболеваемост. ICD-10 е 10-та ревизия на ICD, медицински класификационен списък от Световната здравна организация (СЗО), който се използва до момента на провеждане на това проучване. Той съдържа кодове за заболявания, признаци и симптоми, необичайни находки, оплаквания, социални обстоятелства и външни причини за нараняване или заболявания [50],[51].

Методи

В това изследване се използва анализ на времеви редове с корелационни методи за анализ на качеството на въздуха и данните за здравето. Статистическите методи в изследването попадат в две категоризации: параметрични и непараметрични. Параметричните методи, като корелационен коефициент на Пиърсън, се предпочитат, когато данните са нормално разпределени или има достатъчно голям обем на измерванията, което позволява да се приеме, че данните следват нормално разпределение. Тези методи са подходящи за оценка на линейната връзка между две непрекъснати променливи. Параметричният корелационен тест на Пиърсън (2.1) се използва за сравняване на двата източника на данни за качеството на въздуха. Той осигурява мярка за линейната връзка между двете непрекъснати променливи (обикновено наричана само коефициент на корелация). Коефициентите на корелация за всяка двойка (x, y) се определят за извършване на оценката и стойностите на x и y съответно се заменят с техните рангове. Резултатите от теста са приложени към коефициент на корелация в диапазона от -1 до 1.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2.1}$$

Непараметричните сравнителни методи, като Spearman's rho и U тест на Mann – Whitney, се използват, когато данните не са нормално разпределени или не е уверен, че може да се приеме, че следват нормално разпределение. Тези методи са подходящи за оценка на ранговата връзка между променливи или за сравнение на разликите между две групи, без да се изисква специфично оценяване на линейната връзка. Изборът в тази дисертация е следният: дескриптивен анализ за индивидуалните вариационни редове; корелационен анализ (непараметричен - Spearman's rho) между двойки променливи; междугрупово сравнене (непараметрично Mann–Whitney U test) между средни стойности.

Моделът, който е разработен в тази задача, е за взаимовръзка на показатели за качество на въздуха и здраве въз основа на данни за фини прахови частици и регистрирани постъпления в болници и Бърза помощ. По съвет на лекари, професионалисти за анализ на здравните данни, в това изследване използваме отместване със забавяне от 1, 2 и 3 дни. Като причината за това е, че след експозицията на ВЗ се изисква време, за да развият ефекти върху здравето [35]. По този начин изследването на приемането в болница и записите от Бърза помощ ще покажат не само кои случаи заболявания търпят промяна поради изменения на качеството на въздуха, но можем да видим и как те се развиват във времето. На Фигура 3.1 е илюстриран методът експозиция-ефект, където дозата е експозицията, умножена по времето.

2.4 Резултати

В едно скорошно проучване се предполага, че евтините сензори могат да предложат груби детайли за установеното качество на въздуха, но не са готови за приложения, включващи висока точност [33]. В тази глава се проверява хипотезата, че евтините сензори за качество на въздуха на Luftdaten могат да бъдат добро допълнение към официалните станции за качество на въздуха на ИАОС. Ако изчистим и обобщим данните, можем да имаме корелация, която ще бъде полезна за разглеждане на някои груби детайли. При сравнението сме изчистили данните, когато влажността на въздуха е била над 70%. Тези условия не означават непременно, че данните са неверни, но при такива условия производителят не гарантира прогнозираната точност от 10%.

Резултатите от дневни концентрации за ФПЧ10 и ФПЧ2,5 са обобщени в таблица 2.2 . Средната дневна концентрация на ФПЧ 2,5 е 11,7 g/m3 с диапазон от 2 до 136 g/m3. През периода на изследването 13,2% от дневните концентрации на ФПЧ10 или ФПЧ2,5 не са отговаряли на нормите на СЗО за качество на въздуха (50 g/m3 за ФПЧ10 и 25 g/m3 за ФПЧ2,5). В данните от ИАОС имаше липси за няколко часа до няколко дни (проблеми със захранване, повреди, профилактики), а някои от източниците за болнични прегледи бяха с липсващи данни през почивните дни (събота, неделя, официални празници), поради което беше предприет анализ на пропуските, за да се изключи възможността за отклонение.

Таблица 2.2: Дневната концентрация на ФПЧ10 и ФПЧ2,5, разделена на работни и неработни дни

Ден от седмицата	**N**	**Средно**	**Медиан**	**Мин**	**Макс**	**Персентили**
ФПЧ10 Работни дни	522	34.17	27.5	5	336	20
ФПЧ10 Неработни	230	37.1	26	6	290	18
ФПЧ2,5 Работни дни	522	11.15	9	2	104	4
ФПЧ2,5 Неработни	230	12.93	8	2	113	3

В данните за хоспитализирани пациенти с по-тежко състояние наблюдаваме следните корелации (Таблици 2.3 и 2.4). Констатира се увеличение (спрямо фоновите нива) на респираторни и белодробни заболявания от 1-ви до 3-ти ден след превишение на ФПЧ10 със 120%, както и сърдечна недостатъчност в същия период от

Таблица 2.3: Сравнение на непараметрични данни от хоспитализирани пациенти в дни с превишения на нормите на СЗО за ФПЧ10

Забол.	ICD-10	без отм.		отм. 1 д.		отм. 2 д.		отм. 3 д.	
		Z	p	Z	p	Z	p	Z	p
Респ. сист.	C30-C39	-1,504	0,133	-2,211	0,027	-2,870	0,004	-2,862	0,004
Сърд. удар	I50	-1,729	0,084	-3,656	0,001	-3,475	0,001	-2,821	0,005

Таблица 2.4: Сравнение на непараметрични данни от хоспитализирани пациенти в тежки състояния в дни с превишения на нормите на СЗО за ФПЧ2,5

Забол.	ICD-10	без отм.		отм. 1 ден		отм. 2 дни		отм. 3 дни	
		Z	p	Z	p	Z	p	Z	p
Респ. сист.	C30-C39	-0,361	0,718	-1,270	0,204	-1,265	0,206	-2,119	0,034
Бел. емб.	I26	-1,601	0,109	-2,302	0,021	-2,894	0,004	-1,852	0,064
Сърд. удар	I50	-1,586	0,113	-2,831	0,005	-3,195	0,001	-2,799	0,005

време с 19%. Превишението на ФПЧ2,5 се свързва с 59% увеличение на честотата на белодробна емболия на дни 2 и 3.

Резултатите показват увеличение на контактите с Бърза помощ с 11% в дни с превишаване на ФПЧ10 и 13,5% в дни с превишение на ФПЧ2,5, за период от най-малко 3 дни. Средно увеличение (спрямо фоновите нива) на неврозите на ден 2 след превишаване на ФПЧ10 е 1%, при инфаркти на ден 3 след превишаване е 8%, при инсулти веднага след превишаване с 9%, продължаващи до ден 3 след превишение. Същото важи и за хипертоничните състояния, които нарастват с 5% непосредствено около превишението и до третия ден след събитието.

От анализираните данни се забелязва значителна корелацията на сърдечносъдовите и мозъчносъдовите заболявания, като напълно съвпада с бързото нарастване на риска при ниски нива на фракции на замърсяване с ФПЧ2,5, съвпадаща с подобни открития в други подобни проучвания [75], [64], [70].

Но най-висока корелаия се забелязва при острите инфекции на горните дихателни пътища се увеличават съответно с 47% и в частност пневмониите с 60%. Хроничната обструктивна белодробна болест (ХОББ) след ден е с 36% повече. Що се отнася до астмата, записите за алергична астма нарастват повече в дни с повишено замърсяване в сравнение с неалергичната астма. Амбулаторните прегледи в болниците за леки състояния, които не изискват хоспитализация, логично откриваме повишено присъствие на респираторни диагнози (таблица 2.5).

2.5 Заключение

В резултат на проведените анализи и тестове се доказа, че острата заболеваемост се увеличава в дните, когато концентрациите на фините прахови частици са над нормата на СЗО за качество на въздуха. Това се установи сравнявайки данни за постъпления в болници и случаите на Бърза помощ в дни със и без превишенията на нормите за ФПЧ в София.

Респираторните заболявания показват по-високо и по-бързо развитие, когато замърсяването на въздуха надхвърли здравословните граници. Други заболявания,

Таблица 2.5: Сравнение на непараметрични данни от амбулаторни прегледи в болници с нормите на СЗО за среднодневна концентрация на ФПЧ10

Забол.	ICD-10	без отм.		отм. 1 ден		отм. 2 дни		отм. 3 дни	
		Z	p	Z	p	Z	p	Z	p
Дихателна система	C30-C39	−1.515	0.130	−2.950	0.003	−2.941	0.003	−2.423	0.015
Ушни заболявания	H65,H81	−0.754	0.451	−0.976	0.329	−1.580	0.114	−2.322	0.020
Миокарден инфаркт	I20-I25	−0.244	0.807	−1.393	0.164	−1.876	0.061	−2.104	0.035
Сърдечна недост.	I50	−0.771	0.440	−2.840	0.005	−2.559	0.010	−2.972	0.003
Горни дих. пътища	J00-J06	−3.218	0.001	−3.507	0.000	−3.910	0.000	−4.605	0.000
Бактериална пневм.	J13-J18	−3.171	0.002	−3.395	0.001	−2.556	0.011	−3.683	0.000
Остър бронхит	J20-J21	−3.096	0.002	−2.848	0.004	−1.606	0.108	−1.588	0.112
Долни дих. пътища	J40-J47	−3.988	0.000	−4.049	0.000	−2.842	0.004	−3.481	0.001
Инфекции на кожата	L00-L08	−2.552	0.011	−1.379	0.168	−0.267	0.789	−0.489	0.625
Микоплазмена пневм.	J20.0	−2.330	0.020	−2.327	0.020	−0.430	0.667	−0.431	0.667
Остър бронхит	J20.9	−2.496	0.013	−1.449	0.147	−0.039	0.969	−1.292	0.197
Астма, алерг.	J45.0	−2.426	0.015	−2.808	0.005	−3.577	0.000	−3.643	0.000
Ангина	I20.8	−1.971	0.049	−0.763	0.445	−0.773	0.440	−0.674	0.500

свързани със замърсяването на въздуха, като сърдечно-съдови и мозъчно-съдови, също се увеличават.

Това е сред първите подобни проучвания в България и може да бъде полезно за специалистите в сферата на здравеопазването, учените по околната среда и политиците.

Методът на ефекта на експозиция е сложен и прецизните измервания са много трудни за постигане дори в експозиционна камера. С помощта на методите на епидемиологичните изследвания и статистиката на населението обаче става ясно, че замърсяването на въздуха допринася за очакваните промени в някои показатели на заболяванията. В заключение, подобно на данните от публикувани проучвания в градове в Европа, Америка и Азия, повишените нива на замърсяване на въздуха се свързват и водят до по-високи нива на диагностицирани заболявания. С малки изключения, краткосрочните ефекти върху здравето се измерват чрез използване на осреднени концентрации на замърсяване на въздуха в целия град за показатели за експозиция. Това може да доведе до погрешна класификация на експозицията и по този начин до отклонение [40].

Анализът на данни за качеството на въздуха показа, че използването на по-обширни данни от мрежи с евтини лазерни сензори, обогатява информацията, предоставяна от високо-точните официални станции на ИАОС. Освен, че станциите на ИАОС са малко на брой, те често имат и липсващи данни за дадени периоди. Допълването на данните от официални станции с данни от гражданските мрежи, води до подобрени измервания на връзките между здравето и замърсяването на въздуха, като осигурява по-обстойна експозиция конкретно на ФПЧ2,5.

Глава 3

Калибриране на граждански сензори чрез машинно обучение

За да разберем как можем да подобрим данните, които получаваме от станциите с лазерни сензори за ФПЧ, ще трябва да измерим тяхната грешка в реални условия. За целта поставяме станции с лазерен сензор в непосредствена близост до станциите на МОСВ за измерване на ФПЧ. Сравняваме техните данни и правим калибровъчен модел на база машинно обучение, който подобрява данните от евтините лазерни станции.

Традиционните станции за наблюдение на качеството на въздуха, оборудвани със сложни и скъпи инструменти, са основният източник на данни за оценка на нивата на замърсяване на въздуха. Високата им цена, ограниченото им пространствено покритие и логистичните предизвикателства обаче доведоха до появата на евтини сензори за качество на въздуха като алтернативно решение за наблюдение. Тези сензори предлагат потенциал за разширяване на мрежите за наблюдение на качеството на въздуха, като предоставят данни с по-висока разделителна способност и позволяват по-локални оценки. Евтините лазерни сензори работят на принципа на разсейване на светлината, като откриват и определят количествено концентрацията на ФПЧ във въздуха. Те са много по-рентабилни от трационните станции за измерване, но им липсва надеждност на измерваните данни, поради чувствителността им към температурата и относителната влажност. Тези слабости са особено очевидни, когато става въпрос за високи нива на ФПЧ в комбинация с висока относителна влажност - тогава стойностите измерени за ФПЧ са неестествено завишени. Калибрирането е от съществено значение за осигуряване на точността и последователността на евтините лазерни сензори.

В тази глава представяме нов двустепенен модел за калибриране на евтини лазерни сензори за наблюдение на въздуха, като използваме техники за ML с наблюдение и без наблюдение. Предложеният от нас метод за калибриране има за цел да подобри точността на отделните лазерни сензори, като използва данни от официални станции за наблюдение на качеството на въздуха като референтни. Първата стъпка от нашия модел за калибриране включва контролирана ML, при която обучаваме прогнозен модел, използвайки данни, събрани както от евтини лазерни сензори, така и от референтни станции за наблюдение на качеството на въздуха. Алгоритъмът за контролиран ML научава връзката между показанията на сензорите и съответните референтни измервания, което позволява калибриране на данните от сензорите. Като използваме изчерпателните и точни данни от референтните станции, можем да преодолеем разликата между показанията на евтините сензори и истинските измервания. Втората стъпка на нашия модел използва неконтролирани ML техники за по-нататъшно усъвършенстване на процеса на калибриране. Алгоритмите за неконтролирано обучение, като например групиране и откриване на аномалии, помагат да се идентифицират и коригират отклоненията и несъответствията в данните от сензорите. Като откриваме и отстраняваме нередностите, можем да повишим общата точност и надеждност на евтините лазерни сензори.

3.1 Наблюдение на качеството на въздуха

Това изследване използва пет станции за контрол на качеството на въздуха с традиционни методи за измерване като ръководство. За изчисляване на ФПЧ10 са използвани автоматичните станции за измерване на МОСВ, които използват гравиметрични методи за събиране на проби от ФПЧ върху филтър и претегляне на филтъра, за да се определи масата на събраните ФПЧ. Тези методи са подобни на метода на бета-анализа, но не разчитат на откриване на радиация. Гравиметричните методи са много точни, но изискват специализирано оборудване и обучен персонал за събиране и анализиране на пробите. Измервателните станции използвани в проучването се намират в София, съответно в районите Младост, Дружба, Надежда, Хиподрума и Красно село. Само една от тези станции измерва ФПЧ 2,5 и поради това ФПЧ 2,5 не се използва за справка в това изследване.

Сензорът за ФПЧ, използван в това изследване, е лазерният сензор за частици NovaFitness SDS011. Този тип сензори използват принципа на лазерна дифракция. Лазерът осветява уловените йони, когато улавя разпръснати светлинни вълни под определен ъгъл, докато въздухът преминава през фоточувствителната област на сензора. Създава се континуум на размера на частиците чрез класифициране на тези импулсни сигнали в различни интервали на размера на частиците, за да се измери масовата концентрация на частиците [38].

В това изследване е използвана безжичната сензорна мрежа (WSN) на Luftdaten. Тя е гражданска мрежа и се състои от 300 стационарни сензора, покриващи София. Всеки сензор е инсталиран в пластмасова тръба, която може да се монтира на стени, балкони, стълбове за улично осветление и други конструкции. Разработени са определени насоки, за получаване на най-доброто представяне на емисиите на прахови частици в града с възможно най-малко сензори. WSN използва фиксирани сензори, които са монтирани в мрежа от 1 km, за да се гарантира, че по-голямата част от центъра на града е покрита с подходяща плътност.

3.2 Модел за калибриране на данни

Този раздел описва двустепенния модел на калибриране за евтини лазерени сензори, който използва комбинация от контролирани и неконтролирани техники за машинно обучение (supervised and unsupervised ML). Този модел използва данни от измервания от евтини сензори и стандартни станции за наблюдение на въздуха като референтен инструмент.

Други изследователи доказват, че моделите с корекция на сензорни данни чрез машнно обучение дават добри резултати, когато са обучени спрямо референтен инструмент [16].

3.2.1 Въведение в контролираното машинно обучение

Контролираното машинно обучение е вид изкуствен интелект (ИИ), при който се използват алгоритми за обучение от маркиран набор от данни и за изготвяне на прогнози или решения въз основа на това обучение. Целта на контролираното обучение е да се намери връзката между входните характеристики на набор от данни и съответните им резултати и да се използва тази връзка, за да се правят прогнози за нови, непознати данни.

Алгоритмите за контролирано обучение се обучават върху маркиран набор от данни, който съдържа както входните характеристики, така и съответните изходи. Алгоритмите използват тези данни за обучение, за да научат връзката между входните признаци и изходите и след това използват тази връзка, за да правят прогнози за нови данни. Точността на прогнозите зависи от качеството на данните за обучение и от избора на алгоритъм.

3.2.2 Методология на изследването

Методологията, използвана за калибриране на концентрацията на ФПЧ10 в София, включва няколко стъпки:

1. Почистване на данните: Данните от гражданските станции се пречистват допълнително, чрез ограничаване на нивата на концентрация на ФПЧ10 въз основа на официалните почасови измервания. Станциите с по-малко от предварително зададена прагова стойност (90 дни) също се отстраняват от набора от данни;

2. Корекция на отклоненията: Изследването започва с коригиране на данните от измерванията на граждани, получени от luftdaten. Тези измервания се сравняват с официалните измервателни станции на Изпълнителната агенция по околна среда (ИАОС), за да се отстранят всички инструментални отклонения;

3. Подготовка на данните: Изчисляват се разстоянията между двойките станции и се създават групи от станции в рамките на определено разстояние. За всяка двойка се изчисляват измервания на различията и станциите с високи измервания на различията се отстраняват от набора от данни;

4. Анализ на факторите и характеристиките: Извършва се проучвателен анализ на данните, като се използват официалните данни на ИАОС. Изследва се значението на метеорологичните променливи и инженерните характеристики, които влияят на нивата на замърсяване на въздуха;

5. Модел за калибриране: Изграден е двустъпков модел с машинно обучение, включващ допълнителни фактори от метеорологични параметри и топографски данни. Характеристиките, използвани за калибриране, включват температура, влажност, налягане, концентрация на PM10 от предишния ден, скорост на вятъра и различни фиктивни променливи. Повече информация за двустъпковия метод с машинно обучение ще бъде изложена в тази глава;

6. Оценка и валидиране: Точността на прогнозните модели се оценява с помощта на квадратична грешка. Моделите са сравнени помежду си, както и с първоначалните данни преди калибриране и показват значително подобрение на точността.

3.2.3 Данни и модел

Моделът използва 5 стандартни станции за наблюдение на въздуха, разположени в София и 5 евтини лазерни сензора със спецификациите, показани в предишния раздел. Малкият размер на евтините лазерни станции позволява възможност да бъдат поставени точно до официалните станции на ИАОС.

Входните променливи са относителна влажност (RH), атмосферно налягане (AP), температура (temp). Стойността, за която калибрираме данните, е ФПЧ10. Тези входни данни получаваме, както от стандартни инструменти, така и от евтини сензори. Тъй като времевата разделителна способност на сензора се различава от тази на обикновения инструмент, за изчисляване и оценка е използвана средната стойност за час.

След като температурата, въздушното налягане и влажността влияят върху стойностите на много сензори за качество на въздуха, измерванията на тези променливи често се записват на място и използват в коригиращия модел, както е направено и от други изследователи [31], [80].

3.2.4 Кратко описание на петте използвани ML техники

Тук описваме на кратко петте контролирани техники за машинно обучение, използвани в първата стъпка на модела за калибриране. Тези техники са: линейна регресия (LR), дърво на решенията (DT), дърво на решенията с градиентен тласък (GBDT), случайна гора (RF) и изкуствена невронна мрежа (ANN).

Линейната регресия е прост и широко използван алгоритъм за контролирано обучение, който моделира връзката между зависима променлива и една или повече независими променливи. Дървото на решенията е модел, който използва дървовидна структура на решенията и техните потенциални последици, включително последиците от случайни събития [59]. GBDT е итеративен алгоритъм за дървета на решенията, съставен от множество дървета на решенията, при който решенията на всички дървета се сумират, за да се получи крайният отговор [39]. RF е хибридно дърво за прогнозиране, при което всяко дърво се основава на стойностите на случайна променлива, избрана независимо и с едно и също разпределение за всички дървета в гората [8]. Изкуствените невронни мрежи са математически модел, който симулира поведението на невроните и се променя автоматично чрез обратно разпространение на грешките [22].

3.2.5 Тип на невронната мрежа и трениране на модела

За целите на калибрирането на данните от лазерни сензори за качество на въздуха към данните от официалните станции, е използвана изкуствена невронна мрежа тип многослоен перцептрон (MLP). Този модел е избран, тъй като е подходящ за регресионни задачи, при които се предсказват непрекъснати стойности. Входните данни за модела включват влажност, атмосферно налягане и температура, както и 'сурови' стойности на фините прахови частици (ФПЧ), измерени от нискобюджетни сензори. Целта на мрежата е да предсказва коригирани стойности на ФПЧ, които да съответстват на измерванията от официалните станции за мониторинг на въздуха, използвани като еталон. За тази цел, изходният слой на модела съдържа един неврон с линейна активационна функция, която е подходяща за регресионни задачи, тъй като резултатът е непрекъсната стойност. Архитектурата на модела включва три слоя – два скрити слоя с по 64 и 32 неврона, които използват активационната функция ReLU, и изходен слой с един неврон за предсказване на коригираните стойности на ФПЧ.

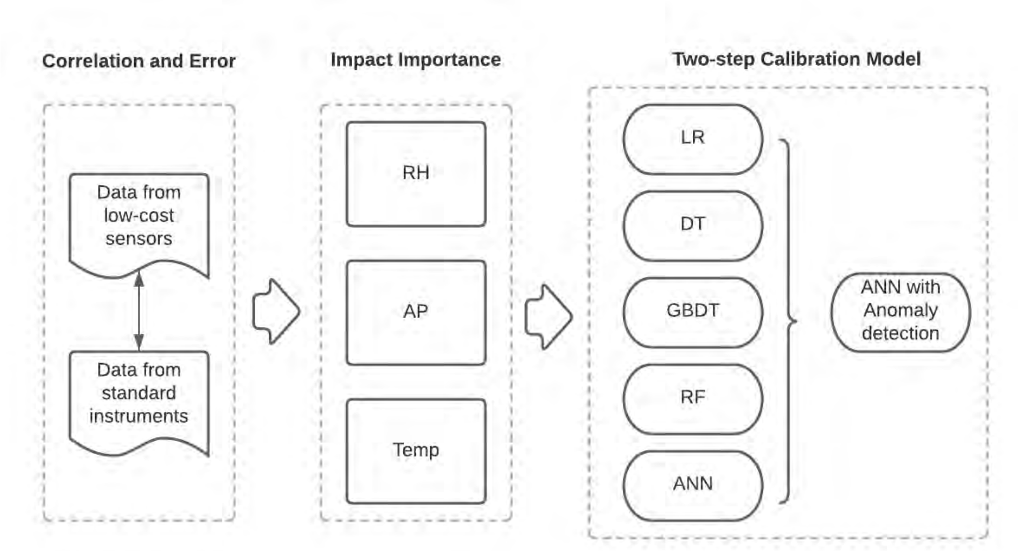
Моделът е обучен върху набор от данни, обхващащ период от три месеца, като входните данни представляват почасови средни стойности на измерванията. Преди обучението всички входни характеристики се нормализират, за да се ускори процесът на конвергенция и да се осигури стабилност на обучението. Като функция на загуба е използван средна квадратична грешка (MSE), която е стандартна при регресионни задачи, тъй като тежко наказва големи отклонения между предсказаните и реалните стойности. За оценка на производителността на модела са използвани метриките R^2, която измерва колко добре моделът обяснява вариациите в данните, и средна абсолютна грешка (MAE), която дава информация за средната разлика между предсказаните и реалните стойности на ФПЧ. Обучението е извършено с помощта на Adam оптимизатор, който е широко използван поради своите адаптивни способности. За да се избегне пренатоварване на модела, е приложена техника за ранно спиране, като процесът е контролиран чрез наблюдение на валидационната загуба, за да се предотврати прекомерното обучение и влошаване на резултатите върху нови данни.

3.2.6 Дизайн на модела

На Фигура 3.1 се демонстрира моделирането на двустъпковия метод за калибриране на данни от станции за наблюдение на възуха. В първия етап на модела се използват 5 контролирани техники за машшинно обучение (supervised ML) и се оценяват техните резултати. Най-ефективният от тези пет медотдa се използва отново във втората стъпка, която включва откриване на аномалии (anomaly detection) за изчистване на отклоненията. На всяка стъпка оценяваме дали това подобрява резултатите на модела и с колко.

Откриването на аномалии и тяхното премахване е втората стъпка в модела и тя може да подобри данните, но това не е сигурно. Това е метод за откриване на необичайни обекти или събития в набори от данни, които са извън обичайното [62]. Детекторът на аномалии е използван за премахване на отклонения само от обучителния набор от данни. Тъй като това е неконтролирано обучение, е направена оценка със същата настройка на ANN, преди и след почистване на набора от данни, за да се определи дали неконтролираното обучение е подходящо за този набор от данни.

Процесът на обучение на модела е разделен на два етапа: обучение и тестване. Необработените данни са разделени на два набора от данни на случаен принцип, с 80% за обучение и 20% за тестване. Моделът първо е обучен с помощта на данните за обучение, а след това изходът му е оценен от тестовия набор от данни.



Фигура 3.1: Двустъков модел на калибриране чрез изкуствена невронна мрежа и намиране на аномалии.

3.3 Приложение на модела върху безжична сензорна мрежа

След калибрирането на 5 сензора, разположени точно до стандартните инструменти, ние прилагаме същия модел към други сензори от мрежата Luftdaten. В анализа са добавени данни от измервания от 13 други сензора, които са в близост до 500 метра до регулаторните станции. Целта на тази корекция е да се прогнозира „реалната" концентрация на ФПЧ10 на място, като се пресъздаде концентрацията на ФПЧ10 възможно най-точно.

За целите на това изследване ние вземаме данни от евтини сензори, които са инсталирани и се експлоатират от гражданите. Поради това има липсващи данни. Процентът на липсващите данни сред евтините сензори е 9,4% в резултат на повредени сензори не са получени данни, като причините за това включват: повреда в работата на кутията (напр. спиране на захранването); проблеми, свързани с времето (няма валидна времева променлива); и невъзможност за предаване на данни. Средният процент пълнота на данните за сензорите е 89,2%, с медиана от 96,7 %.

Обсъдените по-горе пет контролирани ML техники са използвани в първата стъпка от анализа: LR, DT, GBDT, RF модели и ANN. След което е изследвани тяхната ефективност. За оценка на моделите се използва корелационният коефициент за квадратична грешка R^2 като индикатор за ефективност при избора и оценката на нашите модели. По отношение на подбора на променливи, ние избрахме само променливи в етапа на валидиране, които се използват за подобряване на констатациите.

Безжичната сензорна мрежа се състои от сензори, монтирани на различни височини. За оценка на измерването на налягането от евтините сензори и идентифициране на всякакви отклонения използваме барометричната формула и коефициента за корелация на Пиърсън. Като входни данни са използвани както стойностите от

Таблица 3.1: Резултати от супервизирани ML модели

ТИП МОДЕЛ	**MAE**	**MSR**	R^2
LR:	11,19	288,12	0,77
DT:	8,89	170,03	0,86
GBDT:	8,68	145,22	0,89
RF:	7,96	125,57	0,90
ANN:	6,27	83,90	0,94

Таблица 3.2: Важността на параметрите въз основа на RF.

Съдържание	**Параметър**
RH:	46,62%
Температура:	30,71%
Атмосферно налягане:	22,67%

евтини ФПЧ сензори, така и референтни инструменти. Височината на монтаж за всеки евтин сензор е известна и добавена към набора от данни. Следователно можем също така да идентифицираме точно разликата във височината между всеки сензор и да анализираме измерванията за атмосферно налягане.

3.4 Резултати и оценка

В Таблица 3.1 са показани резултатите от петте контролирани ML техники. За по-добра оценка на евтините сензори заедно с коефициента на детерминация (R-квадратично) са изчислени MAE и MSE.

Средната стойност на квадратична грешка между сензорите за ФПЧ и стандартните инструменти без калибриране е 0,62. Моделът с линейна регресия показа най-лошия резултат след калибриране със средна стойност R^2 от 0,77.

Най-добрата корелация за ФПЧ10 е получена от модела с изкуствени невронни мрежи. Средната стойност на R^2 е 0,94 (ФПЧ10), което съвпада с констатациите от предишни проучвания [57], [56].

За дългосрочно сравнение евтиният сензор и обикновеният инструмент са поставени на едно и също място, което е общ подход за оценка на сензорите в предишни проучвания [26], [31], [34].

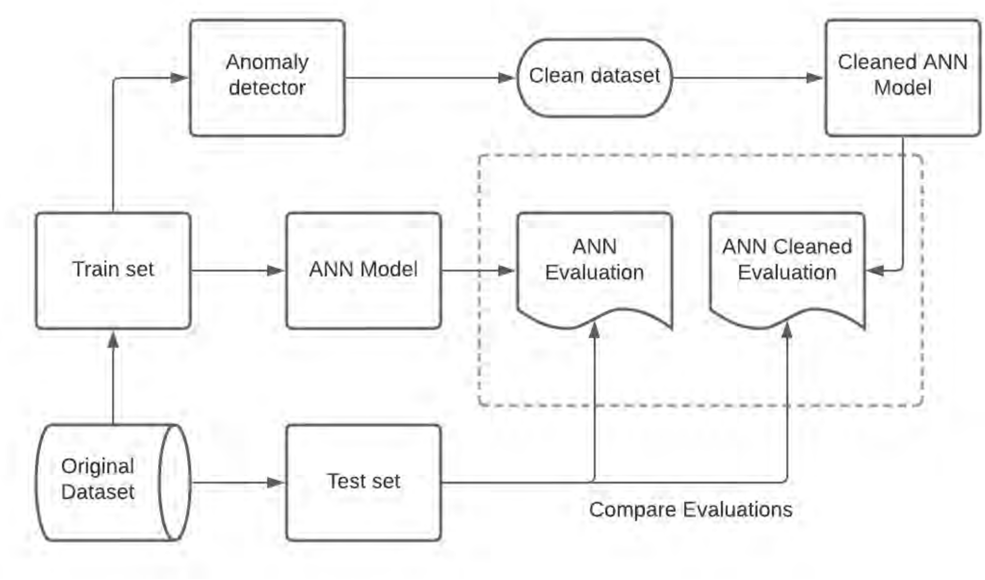
3.4.1 Оценка на резултатите за относителната влажност

В таблица 3.2 се вижда, че RH, последвана от температурата се считат за най-важните фактори, оказващи влияние върху ефективността на сензора за частици. В предишни проучвания е доказано, че високата RH е катализатор за предизвикване на хигроскопичен растеж на частиците и промяна на оптичните свойства, което води до съществени смущения за лазерните сензори за ФПЧ [42]. Това се потвърждава и в моделите ни за RF и ANN за стойностите на ФПЧ10, където RH се оказа с най-голямо значение.

За съжаление, нормалните уреди не могат да бъдат сравнени за съотношението ФПЧ2,5/ФПЧ10, тъй като, както беше споменато по-горе, от петте стандартни уреда

Таблица 3.3: Сравнение на едностъпков и двустъпков модел.

ТИП МОДЕЛ	**MAE**	**MSE**	$\mathbf{R}^2$
едностъпков	6.27	83.90	0.94
двустъпков с откриване на аномалии:	5.62 (↑5.16%)	65.37 (↑14.69%)	0.95



Фигура 3.2: Модел на изкуствени невронни мрежи с откриване на аномалии

само един измерва ФПЧ2,5.

3.4.2 Резултати от модела за калибриране

В Таблица 3.2 са представени статистическите резултати от тестването на всеки модел, където са определени средната абсолютна грешка (MAE), средната квадратична грешка (MSE) и R^2.

Резултатите от останалите пет отделни модела показват, че ANN моделът се е представил най-добре. Моделът RF показа малко по-лоши резултати. R^2 за ФПЧ10 се увеличи от 0,62 до 0,9 и 0,94 съответно за RF и ANN. Моделът ANN се представи най-добре от 5-те модела, малко по-добре от модела RF, поради което беше избран да бъде използван за втората стъпка - с откриване на аномалии. Следователно използването на неконтролирано обучение в това проучване се счита за полезно и подобрява резултата от калибрирането.

В заключение, окончателният модел на ANN (ANN с откриване на аномалии) има най-добър резултат от калибрирането, с най-малка грешка и най-добра корелация, което показва, че предложеният двустепенен модел е по-точен от единичния модел при калибрирането на модела на евтини сензори.

3.5 Заключение и бъдещи изследвания

Ефективността на сензорите за ФПЧ беше измерена чрез сравняване на стандартни инструменти с помощта на безжичната сензорна мрежа. За калибриране на стационарните сензори беше разработен процес в две стъпки и бяха оценени резултатите от модела.

Резултатите от двустъпковия модел са окуражаващи. R^2 на фиксираните сензори за ФПЧ10 се увеличи от 0,62 до 0,95. Моделът ANN имаше най-силно въздействие от петте независими модела, следван от модела RF, докато моделът LR е неефективен.

Детекторите на аномалии могат да бъдат неконтролирана алтернатива на класификаторите при небалансирана съвкупност от данни. В резултат на тях се открива нежелано поведение на сензора и се отстранява от набора от данни. Откриването на аномалии подобри крайните резултати в това изследване.

Относителната влажност се оказа най-важният фактор за резултатите от калибрирането. Това беше нещо очаквано, тъй като високата влажност е условието, при което евтините сензори показват най-много слабости в качеството на данните. Това се оказа така и в това изследване, като имаше по-висок фактор на въздействие от температурата и атмосферното налягане.

Стойностите на атмосферното налягане на стандартните станции и сензори бяха оценени чрез използване на изчисления с барометричната формула. Корелацията беше силна, което означава, че евтините сензори могат да се разглеждат като добър източник за моделиране на замърсяването на въздуха при вертикално планиране в по-нататъшни изследвания.

Този тип калибриране, използвайки контролирана и неконтролирана ML, демонстрира възможността за подобряване на резултатите от евтините сензори. Освен това то може да се използва и за оценка на отклоненията. Могат да се проведат допълнителни изследвания, за да се разбере дали тези отклонения са повредени сензори, неправилно инсталирани такива или те работят добре, но съществуват хиперлокални промени в условията на атмосферата и качеството на въздуха.

Глава 4

Разработка на софтуерен модел за избор на велосипедно трасе с оглед замърсяването от ФПЧ

По-голямата част от специализираните и обособени велосипедни алеи в градовете са в близост до пътни артерии с интензивен трафик, което може да доведе до значителен риск за здравето на велосипедистите поради високия прием на замърсители, свързан с по-високите нива на вентилация [81], [27] и високите нива на физическа активност [7], [53]. Изследователите са се съсредоточили върху оценката на действителните нива на експозиция на велосипедистите по предварително избрани маршрути с помощта на персонални пробовземачи [17] или върху извеждането на личната експозиция от измерванията на замърсяването на улично ниво с помощта на мобилни лаборатории или стационарни станции за наблюдение на атмосферния въздух [52]. Много проучвания също така се опитват да свържат определени физиологични реакции с излагането на велосипедистите на замърсяване на въздуха и откриват доказателства, че краткосрочното излагане може да доведе до вредни последици за здравето [71]. Едно изследване дори установява, че велосипедистите абсорбират по-голяма част от фините частици ФПЧ2,5 и черния въглерод, отколкото водачите на моторизирани видове транспорт [30].

4.1 Формулиране на задачата

Тук е разгледан въпросът за недостигът на количествено определяне на експозицията на градските велосипедисти на ФПЧ в райони с лошо качество на въздуха. В град София, известен с високите си нива на замърсяване на въздуха, липсва цялостен метод за оценка и количествено определяне на нивото на експозиция на велосипедистите. Този пропуск в знанията възпрепятства разработването на ефективни стратегии за минимизиране на експозицията и осигуряване на по-здравословни велосипедни маршрути.

За да се справим с този проблем, предлагаме разработване и оценка на софтуерен модел, който използва модифициран метод за определяне на маршрута, който отразява очакваното вдишване на ФПЧ от велосипедиста, осигурявайки по-благоприятен за здравето маршрут. Чрез включването на данни за качеството на въздуха, концентрациите на замърсителите и моделите на движение в софтуерния модел можем да определим количествено излагането на велосипедистите на ФПЧ и да предложим препоръки за маршрути с минимално излагане. Този новаторски подход има за цел да запълни съществуващата празнина в количественото определяне на експозицията и да предостави практически инструмент за градските велосипедисти в замърсени градове като София.

Софтуерът реализира системата за вземане на решения, която е изложена в тази глава. Той е разгърнат в облачна инфраструктура, а за разработването му са използвани Python, Django и GraphQL (за базата данни). Той комбинира gis данни с данни за трафика и качеството на въздуха, за да приложи модифицирания алгоритъм $k - SPwLO$ за намиране на най-подходящия велосипеден маршрут и да го изобрази на картата $Openstreet$.

Входните параметри са: измервания на ФПЧ от фиксирани сензори; измервания на ФПЧ от мобилни сензори; данни за трафика; температура и влажност от стационарни сензори; начална и крайна точка на пътуването; денивелация на терена; сърдечен ритъм.

4.2 Проект на изследването

В това проучване се разглежда създаването на софтуерен инструмент, който има за цел да избере оптимален маршрут за колоездене, който осигурява най-малката доза вдишване на ФПЧ за велосипедист, който се опитва да стигне от точка А до точка Б.

Инхалационната доза (ID) на замърсителите на въздуха е променлива величина, която зависи от концентрациите на замърсителите, времето и степента на вентилация (VR) [min]. Изчисляваме инхалационната доза, като включваме в модела експозицията на ФПЧ за всеки велосипедист с биомаркери, като сърдечна честота, и времето, необходимо за изминаване на всеки маршрут. Следващият подраздел ще предостави по-задълбочена информация начините на изчисляване, използвани в модела.

За да изчислим VR на велосипедиста (в l/min), използваме уравнението на модела [19], което се основава на честотата на сърдечния ритъм (HR) [min] (уравнение 4.1).

$$VR = 0,00070724 \times HR^{2,17008537} \tag{4.1}$$

За да определим количествата прахови частици, които въздействат на велосипедистите, използваме уравнение 4.2 [19], за да изчислим дозата на вдишване на ФПЧ за всеки участък:

$$\text{ФПЧ}_{inh} = \text{ФПЧ}_{conc} \times VR \times time \tag{4.2}$$

където ФПЧ$_{inh}$ (μg) е масата на замърсителите, които попадат в дихателните пътища на велосипедистите по време на пътуването (в двете посоки); ФПЧ$_{conc}$ (μg$/m^3$) е средната експозиция на замърсителите.

Формулите за изчисляване на инхалационната доза (уравнения 4.1 и 4.2) водят до следната хипотеза: ако искаме да създадем инструмент, който намалява инхалационната доза на ФПЧ, той трябва да избере трасе, което е бързо и кратко. Колкото по-малко е времето, толкова по-малко е ФПЧ$_{inh}$; изисква по-малко усилия. HR се увеличава по време на изкачване и висока скорост. Търсете маршрути с ниска денивелация. Преминава се през по-малки концентрации на ФПЧ; пътищата с интензивно движение трябва да се пропускат. Предпочитат се малки улици и алеи в паркове.

Системата за вземане на решения се калибрира, като се използват данни за HR, IR, скорост на колоездене, измервания на ФПЧ от мобилни и стационарни сензори и данни за трафика. Софтуерът обединява данните за трафика и качеството на въздуха, взети от няколко източника, със специфични данни за велосипедния маршрут, като точното местоположение, времетраене на каране и денивелацията.

За да се изчисли най-малкото вдишване на мръсен въздух, от този софтуер се изисква преди всичко да изчисли най-краткия път в линейно време [4], [72]. Освен това той трябва да представя варианти, които са по-дълги от най-краткия път, но имат различни желани характеристики, като например по-малък трафик на превозни средства и малка денивелация. Проблемът за k-кратките пътища е прост метод за изчисляване на алтернативни маршрути [46].

Ето защо в това проучване решихме да използваме алтернативен маршрут, и по-специално k-Shortest Paths with Limited Overlap (k-SPwLO), който е представен порано в [12]. Приложението k-SPwLO търси пътища, които са: възможно най-кратки; достатъчно различни един от друг.

Въпреки че методът се представя по-добре от базовото решение, което изброява пътищата по реда на нарастващата им дължина, OnePass не е полезен дори за средни по размер пътни мрежи [13]. За тази цел използваме MultiPass - по-точен метод, който чрез добавяне на втори критерии за подрязване, разширява и подобрява OnePass. MultiPass обхожда мрежата k-1 пъти, но оценява и разширява само най-обещаващите пътища, за разлика от OnePass, който обхожда пътната мрежа веднъж и разширява само онези пътища, които отговарят на критерия за сходство. Подрязването се извършва за всеки път, който не може да доведе до решение.

Нека P е съвкупност от маршрути в пътната мрежа G, които свързват възлите s и t. Път $p'(s \to t)$ се нарича "алтернатива"в P, когато p' е достатъчно различен от всеки път $p \in P$. Формално, коефициентът на припокриване между p' и p определя доколко те са подобни:

$$Sim(p', p) = \frac{\sum_{(n_x, n_y) \in p' \cap p} w_{xy}}{l(p)}, \tag{4.3}$$

където $p' \cap p$ показва групата ребра, които са общи за p' и p. Като се има предвид прагът на сходство θ, маршрутът p' е алтернативен на множеството P, ако $Sim(p', p) \leq \theta$.

При даден изходен възел s и целеви възел t заявката k-SPwLO предоставя колекция от k маршрута от s до t, подредени в нарастващ ред по дължина, така че: най-краткият маршрут $p'(s \to t)$ винаги е включен; всички k маршрута са двойно несходни по отношение на прага на сходство θ;

всички k маршрута са възможно най-кратки.

Крайният резултат от функцията k-SPwLO-modified е единичен път. Този път представлява най-краткия път от началния връх до крайния връх, като се вземат предвид определени ограничения, като избягване на големи булеварди и локации с високи нива на ФПЧ2,5 или голям трафик. Алгоритъмът преминава през съседите на всеки връх, като разглежда техните подходящи характеристики, базирани на зададените критерии, и избира пътя с най-ниска цена. Ако алгоритъмът успее да достигне до крайния връх, връща този единичен път.

4.2.1 Измервания от фиксирани сензори

Софтуерът използва инструмент за агрегиране, който извлича данни за качеството на въздуха както от стандартни инструменти за мониторинг на околната среда на ИАОС, така и от евтини сензорни мрежи като Luftdaten, Smog, Openaqi и много други. Той получава, записва, почиства и калибрира данните за качеството на въздуха от фиксирани евтини сензори.

За да се намалят гореспоменатите недостатъци на фиксираните евтини сензори, ние калибрираме получените от тях данни. Това се прави, като се изследват както данните от станциите за наблюдение на околната среда, така и от мрежата от евтини сензори Luftdaten. За да калибрираме данните от евтините сензори и да повишим тяхната надеждност, използваме двустъпковия метод за калибриране [76], който използва изкуствени невронни мрежи и откриване на аномалии.

4.2.2 Измерване на експозицията на ФПЧ посредством мобилни сензори

Евтините преносими сензори за измерване на замърсяване са устройства за наблюдение на околната среда, които хората могат да носят или носят, докато извършват ежедневните си дейности. Тъй като те откриват нивата на замърсяване директно и в реално време, те могат да позволят на доставчиците на здравни услуги и изследователите да наблюдават експозицията на индивидуално ниво и да дадат възможност на гражданите да управляват личната си експозиция на замърсители отвъд това, което регулаторните станции за наблюдение могат да направят [73].

Сензорът за частици в Airbeam2 - PMS 7003 е седмо поколение от серията PMSx003, разработена от Plantower. PMS7003 е цифров и универсален сензор за концентрация на частици, който се използва за измерване на количеството суспендирани частици във въздуха и извеждане на резултатите чрез цифров интерфейс. Той може да се използва с различни инструменти, за да предоставя точни данни за концентрацията в реално време. Осигурява солидна ефективност на броене на частици от 98% и може да работи при температурен диапазон между -10 и 60 C..

4.2.3 ФПЧ измервания и калибриране на модела за вземане на решения

Както вече беше обсъдено, софтуерът е програмиран да търси маршрут, за който: най-краткият маршрут $p'(s \to t)$ винаги е включен; маршрута са двойно несходни по отношение на прага на сходство θ; маршрута са възможно най-кратки.

Изчислителният модел се основава на споменатите по-рано формули за намиране на скоростта на вентилация и инхалация, които са използвани за разработване на теглата за алгоритъма за търсене.

Алгоритъмът е допълнително усъвършенстван чрез оценка на маршрутите, комбинираща данни за трафика заедно с данни за качеството на въздуха от мобилни и стационарни сензори. Обобщаването помага за по-прецизно прогнозиране за изчисляване на най-добрия маршрут. Стационарните сензори, които са разположени в близост до велосипедните маршрути, изготвят базовата линия на концентрациите на ФПЧ в района - $PM_{baseline}$. В същото време данните за трафика дават информация, която помага да се оцени концентрацията на ФПЧ от превозните средства и транспорта - PM_{trafic}. Очакваната експозиция на ФПЧ PM_{exp} по велосипедния маршрут се намира чрез комбиниране на двете променливи.

$$PM_{exp} = PM_{baseline} + PM_{trafic} \tag{4.4}$$

Колкото по-близо минава велосипедният маршрут до главен път, толкова по-голямо е влиянието на замърсяването от трафика върху общите концентрации на ФПЧ. Следователно променливата PM_{trafic} се увеличава. Времето, в което се предприема пътуването, също влияе върху тази променлива - преминаването по време на задръствания и в близост до силно натоварени улици и булеварди увеличава концентрациите на ФПЧ, както е отбелязано по-нататък в главата. Действителните концентрации на ФПЧ се оценяват с помощта на измервания на ФПЧ от мобилни сензори. Мобилните сензори осигуряват много точни измервания, тъй като са прикрепени към самия велосипед, предават измерванията на ФПЧ на всяка минута чрез мобилен телефон и отразяват ултралокално замърсяване, като например да се намираш непосредствено зад автобус или огромен камион. По-късно в полевите изследвания е показано как се извършва тази оценка.

Особеностите за оценка на велосипедните маршрути са силно зависими от местоположението и особеностите на конкретното място [5]. Ето защо ние извършихме много подробно полево проучване, като изследвахме два велосипедни маршрута, описани в следващия раздел.

4.2.4 Полеви тестове и проучвани маршрути

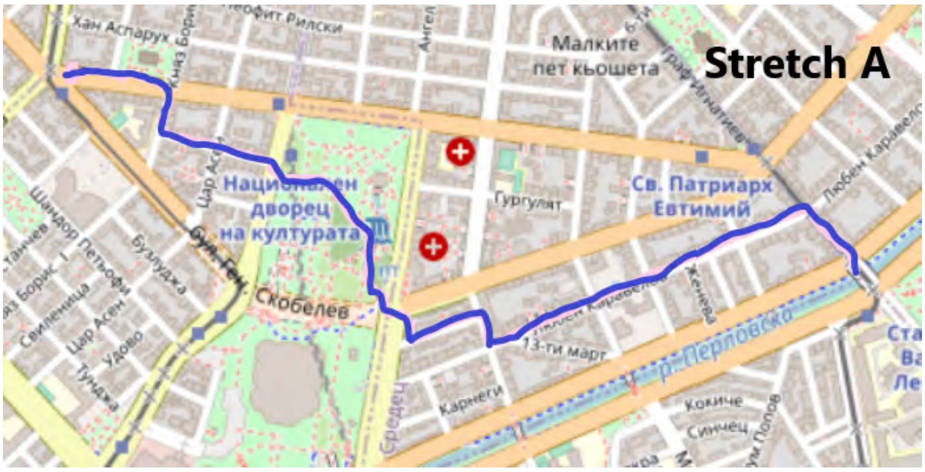
В тази част се разглежда оценката на софтуерният модел с реално теренно проучване в София и разглеждане на два велосипедни маршрута. Двата проучвани маршрута започват от националния стадион "Васил Левски"и завършват в Петте кьошета. Тези две места се намират в центъра на София и имат активен велосипеден поток.

Двата проучвани маршрута са избрани, тъй като отговарят на следните критерии: имат еднакви начална и крайна точка; единият е предложен от софтуера, а другият следва обособени велосипедни алеи; маршрутите трябва да се използват активно от велосипедисти; и двата маршрута минават в близост до стационарни сензори от мрежата Luftdaten; и двата маршрута трябва да са безопасни за редовните велосипедисти.

На следващите изображения се виждат двата пътя, използвани за оценката: Участък А (фиг.4.1), който е оптимизираният велосипеден маршрут, предложен от софтуерния инструмент, и преминава по най-краткия път през малки централни улици и паркове; докато участък Б (фиг.4.2) използва развитата велосипедна инфраструктура (която включва предимно колоездене в близост до пътни артерии с интензивно движение), по-дълъг е като разстояние и е предложен от навигационния софтуер, тъй като следва обособени велосипедни алеи.

Полевите тестове са проведени с 10 участници в изследването. Мобилното оборудване бе свързано към предната част на всеки велосипед, което позволяваше на линиите за вземане на проби да улавят замърсителите, без да бъдат възпрепятствани; то също така беше укрепено отдолу, за да се намалят вибрациите.

Избрахме 10 души (на възраст 27-41 г.) чрез контакт с местната велосипедна мрежа (8 мъже и 2 жени). Преди да се включат в проучването, участниците попълниха



Фигура 4.1: Отсечка А - предложена от нашия софтуер, който търси най-малката инхалационна доза

предварително скринингово проучване. Критериите за изключване от групата бяха респираторни (включително астма), сърдечносъдови или други хронични заболявания, както и тютюнопушене (настоящо или скорошно). Включени бяха само лица, които пътуват ежедневно с велосипед в София. Тези фактори бяха използвани, за да се намали рискът от увреждане поради непознаване на улиците в София, липса на опит в карането и неблагоприятни остри последици за здравето. Освен това участниците бяха помолени да се въздържат от алкохол и кофеин в продължение на 48 часа преди тестовете.

Кръговите пътувания бяха извършени в работни дни сутрин в часовете на интензивно движение (HT) (8:00-9:30 ч.) и в часовете на слабо движение (LT) (10:30-12:00 ч.) и в неработни дни (NWD): почивни и празнични дни. Резултатите от трите сценария са оценени поотделно в следващия раздел. Тестовете на терен са проведени през април и май, когато времето е умерено и замърсяването на въздуха от битовото отопление не оказва влияние върху данните. Следователно увеличенията на концентрациите на ФПЧ се дължат главно на транспорта. Десетимата участници извършиха кръгови пътувания по двата участъка по време на HT, LT и NWD.

4.3 Резултати и обсъждане

4.3.1 Концентрации на ФПЧ1 и ФПЧ2,5 от мобилни измервания

В таблица 4.1 са показани минималните, максималните и средните (медианни) стойности на концентрациите на ФПЧ1 и ФПЧ2,5 за двата изследвани участъка в работни дни по време на интензивен трафик (HT). Концентрациите на ултрафините частици с диаметър под 1 микрон (ФПЧ1) са измерени в диапазона между 8 и 24 μg /m3 (средно 11 μg /m3) за участък А и между 8 и 41 μg /m3 (средно 14 μg /m3) за участък Б. Концентрациите на фини частици под 2.5 микрона (ФПЧ2,5) са между 12 и 29 μg /m3 (средно 15 μg /m3) за участък А и между 12 и 45 μg /m3 (средно 19

Фигура 4.2: Отсечка Б - предлагана от повечето навигационни софтуери, който следва обособена велосипедна алея

μg /m3) за участък Б.

Таблица 4.1: Концентрации на ФПЧ1 и ФПЧ2,5 от мобилни измервания в работни дни по време на интензивен трафик (HT)

	ФПЧ1			ФПЧ2,5		
	min	max	mean	min	max	mean
Участък А	8	24	11	12	29	15
Участък Б	8	41	14	12	45	19

В таблица 4.2 са представени концентрациите на ФПЧ1 и ФПЧ2,5 за участък А и Б в работни дни по време на слаб трафик (LT). Измерванията от участък А показват по-ниски нива на концентрация както за ФПЧ1, така и за ФПЧ2,5. Концентрациите на ултрафини частици с диаметър под 1 микрон (ФПЧ1) са измерени между 6 и 19 μg /m3 (средно 10 μg /m3) за участък А и между 5 и 34 μg /m3 (средно 12 μg /m3) за участък Б. Концентрациите на фини частици под 2,5 микрона (ФПЧ2,5) са между 10 и 24 μg /m3 (средно 13 μg /m3) за участък А и между 9 и 38 μg /m3 (средно 15 μg /m3) за участък Б.

Таблица 4.2: Концентрации на ФПЧ1 и ФПЧ2,5 от мобилни измервания в работни дни със слаб трафик (LT)

	ФПЧ1			ФПЧ2,5		
	min	max	mean	min	max	mean
Учасък А	6	19	10	10	24	13
Учасък Б	5	34	12	9	38	15

Таблица 4.3 илюстрира минималните, максималните и медианните стойности на

концентрациите на ФПЧ1 и ФПЧ2,5 за двата изследвани участъка през почивните и празничните дни. Измерванията за участък А показват почти същите ниски нива на концентрация за ФПЧ1 и ФПЧ2,5. Концентрациите на ФПЧ1 са между 3 и 7 μg /m3 (средно 4 μg /m3) за участък А и между 3 и 8 μg /m3 (средно 4 μg /m3) за участък Б. Концентрациите на ФПЧ2,5 са между 4 и 11 μg /m3 (средно 7 μg /m3) за участък А и между 5 и 11 μg /m3 (средно 7 μg /m3) за участък Б.

Таблица 4.3: Концентрации на ФПЧ1 и ФПЧ2,5 от мобилни измервания през неработни дни

	ФПЧ1			ФПЧ2,5		
	min	max	mean	min	max	mean
Учасък А	3	7	4	4	11	7
Учасък Б	3	8	4	5	11	7

Въпреки факта, че специализираният велосипеден маршрут е през по-голяма част от времето открит, големият брой автомобили, автобуси и камиони в този коридор е основната причина за повишените концентрации на замърсяване, като по-големият трафик увеличава концентрациите на ФПЧ. В резултат на това измерванията на ФПЧ2,5 са почти еднакви през почивните дни, но с почти 20% по-високи концентрации между двата маршрута през работните дни.

4.3.2 Количество на вентилацията

За да измерим скоростта на вентилация (VR) при дишане, направихме измервания на сърдечния ритъм (HR), кислородното насищане (SpO2) и дихателната честота (RR) на всеки участник в проекта. Средната стойност на всеки параметър има ниска променливост и крайните резултати за VR са представени в Таблица 4.4. Не се наблюдават разлики в тези три фактора в зависимост от това дали маршрутът е изминат в работен или неработен ден, тъй като те не се влияят пряко от трафика или различните нива на излагане на краткосрочно замърсяване на въздуха.

Таблица 4.4: Осреднени показатели от полеви измервания

	Участък А	Участък Б
VR (L/min)	10,14	11,06
HR (удари/min)	82,30	85,66
Денивелация (m)	3	15
Скорост (km/h)	14,3	14,7

Въпреки еднаквите начални и крайни точки, разстоянията са различни и участък А е по-къс - 3,8 км в сравнение с 4,4 км от участък Б и е по-бърз - 15:52 мин в сравнение с 18:04 мин от участък Б. Времето, необходимо за изминаване на двата участъка, е по-скоро постоянно през работните и неработните дни. Специализираният велосипеден маршрут е имал по-дълги и прави коридори с по-малко кръстовища и пресечки. Това е довело до по-високи максимални и средни скорости. Освен това участък Б има по-голяма денивелация - 15 м в сравнение с 3 м в участък А. Всички тези данни водят до повишена скорост на велосипедистите в участък Б - 11,06 /мин в сравнение с 10,14 л/мин за участък А.

В таблица 4.5 са показани резултатите от инхалационната доза на велосипедистите за ФПЧ1 и ФПЧ2,5. Те са изчислени въз основа на средната стойност на експозицията на ФПЧ по време на обиколката, измерена от мобилни сензори, заедно с времето, необходимо за нейното извършване, и скоростта на вентилация за всеки велосипедист.

Таблица 4.5: Доза за вдишване на ФПЧ1 и ФПЧ2,5 по време на обиколките

Период	Замърсител	Участък А (оптимизиран)	Участък Б (велосипедни алеи)
Интензивен трафик	ФПЧ 1_{inh}	29.74	46.45 (↑56%)
	ФПЧ2.5_{inh}	40.56	63.04 (↑55%)
Слаб трафик	ФПЧ 1_{inh}	27,04	39,82 (↑47%)
	ФПЧ2.5_{inh}	35.15	49.77 (↑42%)
Неработни дни	ФПЧ 1_{inh}	10.82	13.27 (↑23%)
	ФПЧ2.5_{inh}	18.93	23.23 (↑23%)

Участък Б показва повишена инхалационна доза на ФПЧ, тъй като завършването на маршрута отнема повече време, изисква повече усилия и велосипедистът е изложен на по-високи концентрации на ФПЧ поради изгорелите газове на превозните средства. Дори и през почивните дни, когато концентрациите на ФПЧ са сходни по двата маршрута - участък Б показва по-високи инхалационни дози поради по-продължителното време и по-високата скорост на вентилация (заради по-висока денивелация, по-висока средна скорост).

Измерванията за SpO2 и RR не показват особени краткосрочни ефекти върху здравето. Това се очакваше, тъй като участниците в проучването, от съображения за безопасност, са подбрани непушачи без хронични заболявания и редовни велосипедисти. Като това не означава задължително, че хората с хронични заболявания и чувствителност към замърсяването на въздуха може да не получат някои симптоми или дразнения, както се наблюдава в някои проучвания [56], [15].

4.3.3 Визуализация на велосипедни маршрути чрез включване на данни от стационарни сензори

Ще използваме мрежата от фиксирани евтини сензори на Luftdaten, за да демонстрираме технологията за нашите фиксирани сензори в нашия софтуер. На фиг.4.3 могат да се видят концентрациите за участък Б по време на LT, където на участъка се прилагат измервания от стационарни сензори в близост до 200 м или по-близо до маршрута. Черно оцветената линия означава, че в тази част от маршрута няма фиксиран сензор, който да е по-близо от 200 м, докато зелените и жълтите цветни линии представят концентрациите, измерени от фиксираните сензори в близост. За стойностите на концентрацията на ФПЧ2,5 между 0 и 12 - се използва зелен цвят, докато жълтият цвят означава стойности на ФПЧ2,5 между 12 и 35. Тези цветови категории са вдъхновени от индекса за качество на въздуха на EPA и са същите за визуализиране на маршрутите на Aircasting, измерени с преносими сензори. Благодарение на гъстата мрежа от евтини сензори в София, за участък А се използват 5 фиксирани сензора, а за участък Б - 6 сензора, тъй като те отговарят на критериите за подбор.

Фигура 4.3: Маршрут за колоездене с експозиция от фиксирани сензори

Софтуерът има постоянен достъп до концентрациите на ФПЧ от фиксираните сензори, а данните за трафика могат динамично да променят входните стойности за механизма за вземане на решения. Инструментът разполага и с функцията да изобразява един и същ маршрут в различни времеви рамки и съответно да показва различни концентрации на замърсяване на въздуха.

4.3.4 Прецизиране на модела и оценка на данните от мобилни и стационарни сензори

Средните стойности за концентрациите на ФПЧ2,5 от стационарните сензори, които са разположени в близост до изследваните маршрути (разстояние 200 м), са сравнени с измерванията на ФПЧ2,5 от мобилните сензори, както е показано в Таблица 4.6 по-долу. Установяваме, че съществува силна корелация между средните стойности на $PM2.5_{fixed}$ и минималните стойности за $PM2.5_{mobile}$, което отчасти доказва предположението, че концентрацията от близките фиксирани сензори може да се използва като изходна за изчисляване на общата оценка на ФПЧ.

Данните от фиксираните сензори са близки до минималните концентрации по маршрута. Обикновено тази част от маршрута е с незначителен трафик и нейната стойност често е близка до $PM_{baseline}$, която се измерва от фиксираните сензори. В същото време максималните концентрации на ФПЧ по велосипедния маршрут се наблюдават, когато се преминава покрай по-значителни замърсители като камиони, превозни средства със стари дизелови двигатели, мотопеди или колони от автомобили, чакащи на светофарите. Както се очакваше, разликата между минималните и максималните концентрации на ФПЧ е по-значителна по време на НТ.

4.4 Заключение

В тази глава е представена разработката на софтуерен инструмент, който оптимизира велосипедните маршрути въз основа на алгоритми, които предвиждат най-малко

Таблица 4.6: Концентрации на ФПЧ2,5 от стационарни сензори, разположени в близост до маршрута, и мобилни измервания

		$2,5_{fixed}$		$2,5_{mobile}$	
Период	Маршрут	mean	min	max	mean
WD-HT	Участък А	10	12	29	15
	Учасък Б	10	12	45	19
WD-LT	Участък А	8	10	24	13
	Участък Б	8	9	38	15
NWD	Участък А	4	4	11	7
	Участък Б	4	5	11	7

вредни замърсители на въздуха. Алгоритъмът използва модифицирана реализация на метода за намиране на алтернативни маршрути - k-кратките пътища с ограничено припокриване. Той се основава на експериментални данни и уравнения, които изчисляват общата доза вдишани замърсители от велосипедиста. Заедно с това изследването е оценено чрез два велосипедни маршрута: Участък А - предложен от новоразработения софтуер, който преминава през малки улици в София, и участък Б - предложен от навигационни приложения, който преминава по обособена велосипедна алея. В полевия тест десет велосипедисти правят кръгови пътувания по двата маршрута през 3 периода: (1) с интензивен трафик, (2) със слаб трафик в работни дни и (3) в неработни дни. Въз основа на събраните в проучването данни от мобилен сензор са изчислени експозицията на велосипедистите и потенциалната инхалационна доза на замърсяване с ФПЧ1 и ФПЧ2,5 по двата маршрута.

Концентрациите на експозиция по специализираните велосипедни алеи се оказаха по-високи от нивата на експозиция по оптималното трасе, особено в работни дни. Дори в случаите, когато средните концентрации са били почти еднакви, инхалационната доза за велосипедиста винаги е била по-висока по маршрута с велосипедна алея, тъй като той е по-дълъг като време и разстояние, с по-висока денивелация и изисква по-интензивно колоездене. При избора на оптимизирания велосипеден маршрут дозата на вдишване на ФПЧ2,5 се намалява с 23%, в неработни дни до 55% по време на интензивен трафик в работни дни. По време на работните дни резултатите показват допълнително повишаване на риска за здравето на велосипедистите, използващи изследваните велоалеи, поради появяването на замърсители директно свързани с трафика. В изследваните велосипедни алеи излагането на велосипедистите на ФПЧ е било тясно свързано с автомобилния трафик, тъй като проучването не е проведено през отоплителния сезон.

Резултатите от това проучване надграждат предишни констатации [3], които показват, че препроектирането на улици за мултимодален трафик с ниска скорост без бариери е по-устойчив и прагматичен подход от изграждането на велосипедна инфраструктура върху пътни артерии с интензивен трафик. Съществува значителна разлика между изследванията със слаб трафик и с интензивен трафик. Велоалеите без физически бариери между велосипедния маршрут и пътя имат по-висока експозиция - заключение, което се установява в друго проучване [37]. Нашите констатации допринасят за по-доброто разбиране на проблемите на София, свързани със замърсяването от мръсен въздух и в частност от автомобилен трафик.

Глава 5

Софтуерна система за визуализация на въздушното замърсяване

За да можем да използваме всички описани до този момент изследвания и да предадем по достъпен начин информация как замърсяването на въздуха влияе върху здравето, е необходимо да създадем комплексна софтуерна система, която позволява събирането, обработката и моделирането на данни.

Архитектурата на софтуера е модуларна и тя определя начина, по който различните компоненти взаимодействат, как се обработват данните и как се визуализират резултатите. На фигура 5.1 е представена как изглежда тя. Архитектурата на софтуера се състои от следните компоненти: модул за свързване с API; Централен репозиториум в облака; модул за обработка на данни; модул за прогнозиране; модул за измерване на замърсяване от ФПЧ по зададено трасе; визуализация и разпространение на данни.

Фигура 5.1: Софтуерна архитектура

Осигуряването на точността и надеждността на нашата система за вземане на решения е от първостепенно значение за успеха на нашето изследване. За да постигнем това, нашият софтуер разчита на редица входни параметри. Тези параметри включват: измервания на ФПЧ от стационарни лазерни сензори; измервания на ФПЧ от официални станции за измерване; измервания на прахови частици ФПЧ от преносими сензори; данни за температурата и влажността от стационарни сензори; данни за трафика.

Всеки от тези входни параметри допринася за по-цялостно разбиране на качеството на въздуха. Комбинацията на данни от системи, които са разделени и не си комуникират по между си в една обща, позволява на нашата система да прави информирани прогнози и оценки на заърсяването и съответно да взима по-надеждни решения.

Изходните данни включват: данни за качеството на въздуха спрямо локация, географски радиус или град в реално време; разположение на станциите с техните стойности върху географски карти; прогноза за качество на въздуха; калибрирани данни от определени станции; избор на най-подходящ път спрямо замърсяване от ФПЧ и автомобилен трафик.

5.1 Модул Събиране на данни

Данните от официални и граждански мрежи със станции за измерване на качеството на въздуха се достъпват посредством заявки към API-та (Luftdaten, aqicn, airthings, openaq и др.). Като системата прави тези заявки през различни периоди. Тези данни се получават в различен формат и времева рамка, за което е нужна допълнителна обработка от системата.

- Замърсители във въздуха: Получаваме информация за различни замърсители във въздуха, включително частици (ФПЧ2.5 и ФПЧ10), азотен диоксид (NO2), серен диоксид (SO2) и въглероден оксид (CO). Всеки замърсител допринася за общия индекс на качеството на въздуха.

- Данни за трафик по пътищата: Информация за трафика по пътищата в реално време ни помага да разберем как емисиите от автомобилите се влияят на качеството на въздуха. Този източник на данни ни помага да идентифицираме точките със задръствания и времевото изменение на емисиите.

- Температура: Знанието за температурата е от съществено значение, тъй като влияе на разпространението и концентрацията на замърсители във въздуха. Често по време на зимния отоплителен сезон се наблюдават по-високи нива на замърсяване, както и температурни инверсии.

- Влажност: Нивата на влажност влияят на качеството на въздуха, тъй като оказват влияние върху концентрацията и разпространението на частиците и замърсителите във въздуха. Високата влажност може да подобри качеството на въздуха, докато ниската влажност може да засилва проблемите обвързани със запрашаване. Влажността е също важен фактор при сравняване и обработка на данни за замърсяване идващи от различни измервателни уреди. Лазерните измерващи станции показват по-високи нива на грешка, когато има висока влажност.

- Налягане на въздуха: Данните за налягането на въздуха помагат при прогнозирането на метеорологични условия, които могат да влияят на качеството на въздуха. Това е важен фактор при моделирането на прогнози за замърсяване на въздуха.

5.2 Модул Съхранение на данни в облака

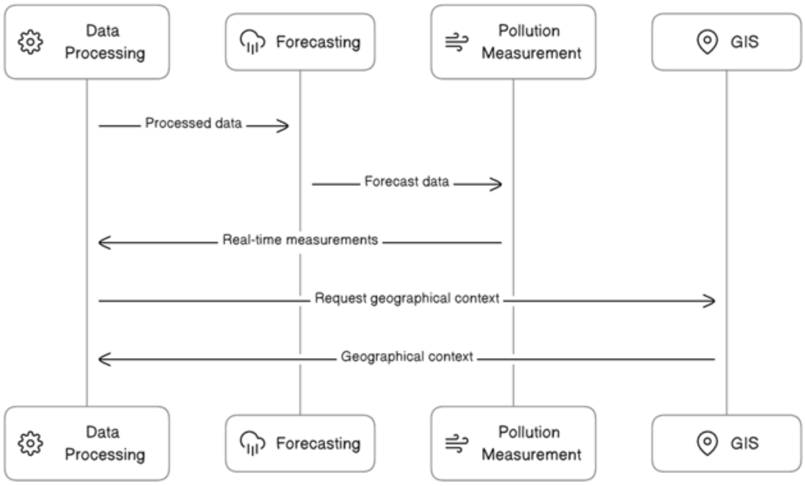
Системата за наблюдение на качеството на въздуха и влиянието му върху здравето на хората се основава на интегриран подход, който включва събиране и обработка на разнообразни данни. Ние събираме данни за замърсители във въздуха, движението по пътищата, температурата, влажността и налягането от различни източници и ги централизираме и съхраняваме в облаковата инфраструктура. Този подход осигурява скалируемост и гъвкавост, позволявайки ни да следим нарастващите изисквания към системата.

Чрез събирането на данни за замърсители във въздуха, движението по пътищата, температурата, влажността и налягането, ние можем да предоставим обстойна представа за качеството на въздуха. Този подход не само засилва нашето разбиране за качеството на въздуха, но и ни позволява да вземем информирани решения и да предприемем своевременни мерки за защита на общественото здраве и околната среда. Данните за трафика са интегрирани, тъй като те предлагат информация за емисиите от превозните средства и тяхното въздействие върху качеството на въздуха. Чрез комбиниране на данните за трафика с измерванията на качеството на въздуха можем да определим областите с високо замърсяване, дължащо се на задръствания, и да разработим стратегии за намаляване на замърсяването.

Данните се актуализират в реално време, което ни позволява да предоставяме актуална информация за качеството на въздуха на потребителите. Интеграцията на данни от различни източници прави анализа и извличането на информация по-лесни и ефективни, предоставяйки обединена платформа за обработка на данни. Потребители с различни права имат лесен достъп до информацията от всяко място с интернет връзка.

5.3 Модул Анализ, прогнозиране и измерване на въздушно замърсяване

Модулът за Обработка на Данни е сърцето на системата и изпълнява няколко ключови функции. Фигура 5.2 предоставя изчерпателен архитектурен преглед на системата за измерване на качеството на въздуха. Тя визуално обрисува сложния поток на данни между модулите за обработка на данни, прогнозиране и измерване на замърсяването. Диаграмата ясно демонстрира как данните се обработват, използват за прогнозиране и се измерват в реално време. Освен това тя подчертава интеграцията на географски контекст чрез данни от ГИС, което улеснява холистичното разбиране на функционалностите на системата. На първо място, този модул предоставя основна предподготовка на данните, включително почистване и структуриране. Това означава, че данните, получени от различни източници и API-та, се подлагат на анализ и обработка, за да се гарантира тяхната точност и цялост. При почистването се извършва отстраняване на невалидни или повредени данни, а структурирането им в подходящ формат осигурява по-лесна навигация и анализ на информацията. Важно е да се подчертае, че модулът включва двустъпков метод за калибриране с машино обучение, който увеличава точността на данните и ги подготвя за анализ.



Фигура 5.2: Архитектурна диаграма, показваща потока на данни между модулите за обработка на данни

5.4 Използвани технологии

Python се използва като основен програмен език за разработка на системата, осигурявайки гъвкавост и мощност в обработката на данни. Неговата лесна четимост и богатият набор от библиотеки го правят подходящ избор за обработка и анализ на данни.

Използването на Python и библиотеките за моделиране носи няколко значителни предимства за нашата софтуерна система. Python предоставя удобна среда за разработка и тестване на модели, като позволява бързо прототипиране и ефективно моделиране на данните.

Python предоставя обширна и гъвкава схема от библиотеки, които са важни инструменти за обработка на данни. Тези библиотеки, като NumPy, pandas и SciPy, предоставят мощни инструменти за анализ, манипулиране и обработка на числови данни. NumPy, например, предоставя възможност за ефективни операции с многомерни масиви и матрици, което е от решаващо значение при работата с данни, получени от физически измервания, където структурата на данните е сложна и изисква бързо и ефективно обхождане.

Построен върху Python, Django предоставя комплексни инструменти за бърза и удобна разработка на уеб базирани приложения. Важен аспект на този инструментариум е вградената обработка на данни и възможността за съхранение и обработка на информация в бази данни. Django предоставя мощна и устойчива рамка за разработка на уеб приложения, подходяща за целите на софтуера.

Съчетаването на Django с инструменти за обработка на данни, като NumPy и pandas, може да бъде изключително полезно при разработка на уеб приложения, които изискват манипулиране и анализ на данни. Интеграцията на тези инструменти помага за създаването на мощни уеб базирани приложения, които работят с данни от различни източници и осигуряват ценни аналитични възможности.

Освен това, Django предоставя възможности за изграждане на публични API, което е изключително полезно за създаването на уеб приложения, които предоставят достъп до данни чрез стандартизирани интерфейси. Това улеснява споделянето и интеграцията на данни между различни системи и приложения.

GraphQL представлява мощен и гъвкав заявителен език за работа с данни, който се използва за взаимодействие с различните видове източници на данни, включително бази данни, уеб услуги и други. Чрез използването на този подход към работа с данни осигуряваме голяма гъвкавост и контрол върху заявките за данни, позволявайки точното извличане на само тези данни, които са необходими за конкретната задача или приложение.

GraphQL не само опростява работата с данни, но и подпомага ефективното взаимодействие между клиентската и сървърната част на приложението, като предоставя оптимизирани заявки и отговори. Този абстрактен и гъвкав подход към работа с данни съчетан с инструменти за обработка на данни, като NumPy и pandas, може да бъде от съществено значение при създаването на приложения, които изискват сложни и динамични заявки към данни.

Използването на GraphQL ни помогна да създаваме приложения, които не само предоставят потребителско удовлетворение, но и улесняват извличането и анализа на данни, което е от съществено значение в днешния свят на информация и данни.

REST (Representational State Transfer) е архитектурен стил за създаване на уеб услуги и API, който предоставя сигурен и ефективен начин за комуникация между различни системи и приложения. REST API предоставят лек и прост начин за достъп до ресурси чрез обичайни HTTP заявки, като GET (за извличане на данни), POST (за създаване на нови данни), PUT (за актуализация на данни) и DELETE (за изтриване на данни).

REST API осигурява лесен начин за достъп и управление на данни чрез стандартизирани и интуитивни заявки, което го прави подходящ за интеграция в различни приложения и платформи. Този структуриран и прозрачен начин на работа с данни осигурява гъвкавост и скалируемост на приложенията.

Облачният репозиториум служи като централно хранилище за данните, идващи от различни източници. Това включва информация за концентрации на замърсители, метеорологични данни, трафик и много други фактори, влияещи върху качеството на въздуха. Облачната инфраструктура осигурява сигурно съхранение и управление на огромни обеми данни.

Използването на облачна инфраструктура е ключово за обработката на големи обеми от данни и предоставянето на услуги, които изискват гъвкавост и скалируемост. Използвани са облачните платформи на Amazon Web Services (AWS) и Microsoft Azure, които предоставят ресурси и услуги за съхранение, обработка и анализ на данни в глобален мащаб. А в комбинация с Python, Django, GraphQL и REST API съвместно с облачната инфраструктура предоставяме инструменти за анализ на качеството на въздуха и влиянието му върху здравето.

5.5 Пример за използване на софтуерната инфраструктура

В основата на функционалността на нашия софтуер стои способността му да интегрира данни от ГИС, измервания на качеството на въздуха от стационарни измервателни станции и данни за трафика. Тази интеграция се постига чрез усъвършенствани техники за сливане на данни, които ни позволяват да получим цялостна представа за качеството на въздуха в различни пространствени и времеви мащаби.

Данните от ГИС осигуряват съществен географски контекст, който ни позволява да определим точно измерванията на качеството на въздуха. Чрез наслагване на данни за качеството на въздуха и трафика върху ГИС слоеве можем да идентифицираме пространствени модели, горещи точки и корелации, които иначе биха останали незабелязани. Тази интеграция е от решаващо значение за нашите изследвания, тъй като ни позволява да оценим въздействието на моделите на движение и географските характеристики върху нивата на качеството на въздуха.

Софтуерът има пособността да визуализира информация за качеството на въздуха върху популярни платформи за картографиране, като Google Maps и OpenStreetMap. Тази функция трансформира необработените данни във формат, който може да се тълкува визуално, като улеснява разпространението на важна информация сред широка аудитория. Можем да анализираме вариациите на качеството на въздуха във времето, като идентифицираме тенденции и аномалии, които да послужат като основа за нашите изследователски резултати. Този времеви анализ дава представа за това как се променя качеството на въздуха през различните часове на деня, сезони или при определени метеорологични условия.

Интегрирането на тези разнообразни източници на данни в съгласувана рамка ни дава възможност да направим значими заключения за качеството на въздуха и взаимодействието му с околната среда, като допринася за дълбочината и всеобхватността на нашите изследвания.

AirLief е мобилно приложение за качество на въздуха, разработено за платформите на iOS и Андроид. То е безплатно приложение с отворен достъп, като показва на потребителя колко е замърсен въздуха около него на база неговата геолокация. AirLief взима данните от софтуерната система, използвайки API-то, създадено за комуникация на данни извън системата.

Данни за състоянието на въздуха и атмосферни данни се разпространяват от софтуерната система към мобилното приложение чрез създадени за целта инфраструктурни крайни точки (ендпойнти). За лесно подаване на данни към нашата система влиза информация за локация (GPS координати), географски радиус или град и излизат данни за качеството на въздуха в реално време спрямо направената заявка. За приложението AirLief са направени 3 инфраструктурни крайни точки (ендпойнти):

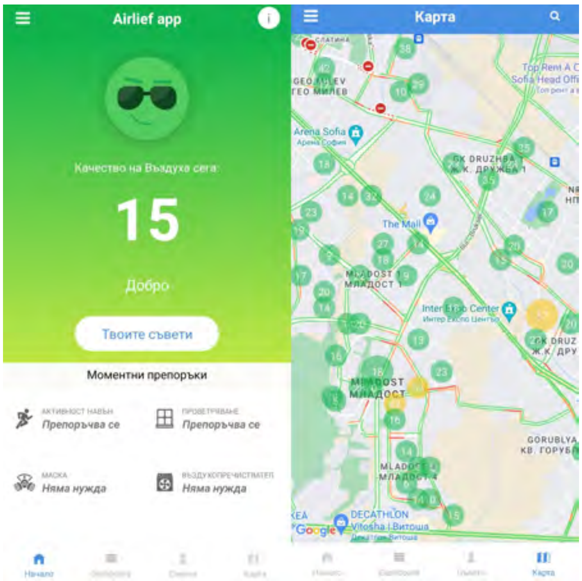
- Измервания с най-близката точка (входни данни: GPS координати): Този ендпойнт позволява потребителите да получат резултати от най-близката измервателна станция, разположена до мястото със зададените GPS координати. Това осигурява информация за текущото качество на въздуха в конкретната локация.

- Измервания по местоположение (входни данни: GPS координати, радиус в км): Този ендпойнт позволява на потребителите да получат средноаритметични стойности на измервания от всички измервателни станции в радиус от определени километри около зададените GPS координати. Това дава обобщена представа за качеството на въздуха в определен район.

- Измервания по град (входни данни: име на града): Този ендпойнт позволява на потребителите да получат средноаритметични стойности на измерванията от всички измервателни станции в определен град или населено място. Това предоставя информация за качеството на въздуха в рамките на конкретния град или населено място.

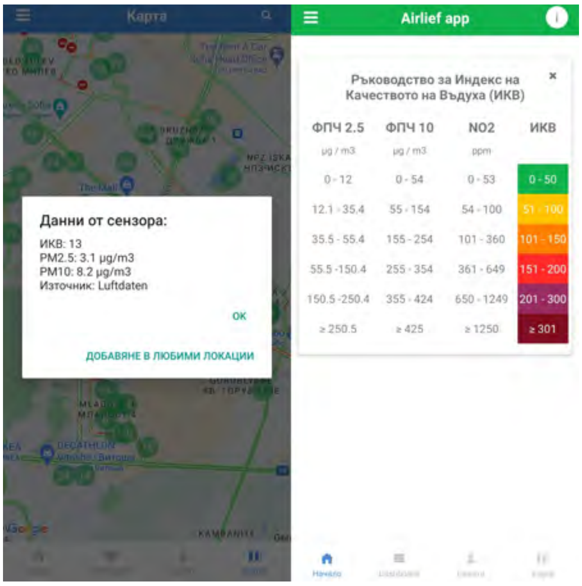
Тези инфраструктурни крайни точки предоставят гъвкавост и удобство при извличане на данни за замърсяването на въздуха и допринасят към обширната функционалност на софтуерната система за следене и управление на качеството на въздуха. В мобилното приложение са показани на потребителите по лесен и достъпен начин.

AirLief app получава данни и препредава по достъпен начин информация за въздушното замърсяване за над 100 държави от около 15000 станции. А самото приложение се използва в 70 държави от над 70000 потребителя, Фигура 5.3. То се състои от дашборд, който показва на потребителя по достъпен и лесен за разбиране начин моментното състояние на въздуха; карта, която показва стойностите от стационарните станции; възможност за доавяне на любими станции и съвети за персонално предпазване от въздушно замърсяване. На Фигура 5.4 е изобразено как самото мобилно приложение превръща суровите данни получени от софтуерната система в Индекс за Качеството на Въздуха (ИКВ). Индекс, който директно корелира с въздействието на въздушното замърсяване върху човешкия организъм, като в него влизат освен измарванията на ФПЧ и други, като озоново замърсяване, азотен диоксид и други.



Фигура 5.3: AirLief App - едно от приложенията на разработената софтуерна система, което показва актуално въздушно замърсяване от стационарни сензори

На фигура 5.4 се вижда как приложението.



Фигура 5.4: AirLief App - мобилното приложение, което превръща суровите данни в разбираем Индекс за Качеството на Въздуха (ИКВ)

Глава 6

Заключение

В дисертационния труд е дефинирана цел и 4 задачи за нейното постигане.

Дисертационният труд изследва връзката между въздушното замърсяване и острите заболявания, базирайки се на данни от две болници и бърза помощ в София. Разработено е решение за калибриране на данни от евтини IoT станции за измерване на ФПЧ чрез машинно обучение, което подобрява точността на измерванията. В тестовете с изкуствени невронни мрежи R^2 се подобрява значително (от 0.62 на 0.95). Изработен е и метод за оптимизация на велосипедни маршрути според нивата на замърсяване, като тестовете показват 67% по-добри резултати спрямо алтернативни решения. Всички задачи са интегрирани в мащабируема софтуерна система за обработка и визуализация на данни, използвана и от мобилно приложение с хиляди потребители.

Научно-приложни приноси

1. Направен е статистически анализ за корелация между въздушното замърсяване и постъпленията в болници и бърза помощ, с които са определени важни причинно-следствени връзки.

2. Разработен е алгоритъм за калибриране на лазерни сензори чрез двустъпков метод с помощта на изкуствени невронни мрежи и детекция на аномалии. Този иновативен метод поддържа по-надеждни и точни измервания от лазерните сензори и има потенциал да бъде използван в широк спектър от сензорни приложения.

3. Разработен е алгоритъм, който оценява влиянието на влажността, височината и атмосферното налягане върху качеството на данните за въздушно замърсяване от лазерни сензори. Това позволява по-пълно и точно разбиране на факторите, влияещи върху замърсяването на въздуха и може да бъде приложим и във вертикално планиране.

4. Разработен е алгоритъм за изчисление на най-подходящ път за преминаване на велосипедист, основавайки се на концентрацията на ФПЧ. Този метод на маршрутизация подпомага здравословния начин на живот и насърчава устойчивия начин на придвижване в градовете.

Приложни приноси

1. Разработен е софтуер за калибриране на данните за ФПЧ от лазерни сензори, взимайки реперни данни от официалните измервателни станции на МОСВ. Този софтуер позволява на организациите за контрол на качеството на въздуха да използват по-надеждни данни за вземане на решения и управление на замърсяването.

2. Разработен е софтуер, който отчита влиянието на влажността, височината и атмосферното налягане и подобрява данните за ФПЧ от лазерни сензори. Този софтуер може да бъде използван от градските власти и здравни организации, за да наблюдават и управляват качеството на въздуха в реално време.

3. Разработен е софтуер за намиране на най-подходящ път за преминаване на велосипедист базирайки се на инхалацията на ФПЧ. Този софтуер може да бъде полезен за индивидуални велосипедисти и градски инфраструктури, които насърчават велосипедното движение.

4. Разработена е IoT платформа за агрегация и анализиране на сензорни данни за качество на въздуха. Този софтуер подпомага автоматизацията и улеснява интеграцията на данни от различни източници, подобрявайки ефективността и точността на наблюдението за качеството на въздуха. Освен метеорологични данни интегрирани са и данни за трафик и GIS. Платформата визуализира карти, като ги допълва с измервания от сензорите за качество на въздуха и данни за настоящо състояние на трафика.

6.1 Публикации по темата на дисертационния труд

- Zhivkov, P., & Simidchiev, A. (2024). Software Tool for Optimizing Cycling Route by Defining Cyclist Air Pollution Exposure
  Studies in Computational Intelligence, 1158 SCI, pp. 152–170, SJR 0.208 (2023)

  https://www.scopus.com/record/display.uri?eid=2-s2.0-85190651936&origin=resultslist&sort=plf-f&src=s&sid=3eca39362dd1584d77f2fb4c5901e3b2&sot=aut&sdt=a&sl=18&s=AU-ID%2857221769514%29&relpos=0&citeCnt=0&searchTerm=

- Zhivkov, P., & Simidchiev, A. (2022). Development of Software Tool for Optimization and Evaluation of Cycling Routes by Characterizing Cyclist Exposure to Air Pollution
  Annals of Computer Science and Information Systems, 32, 105-112.

  https://d1wqtxts1xzle7.cloudfront.net/99417790/230-libre.pdf?1677962458=&response-content-disposition=inline%3B+filename%3DDevelopment_of_Software_Tool_for_Optimiz.pdf&Expires=1725292377&Signature=QB4TSFkNK4~abulkZUncpTN5Hj4TAjM9gME_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

- Zhivkov, P. Optimization and Evaluation of Calibration for Low-cost Air Quality Sensors: Supervised and Unsupervised Machine Learning Models Annals of Computer Science and Information Systems, Volume 25, FedCSIS 2021, 2021, pp. 255–258

  https://www.scopus.com/inward/record.uri?eid=2-s2.0-85117793346&doi=10.15439%2f2021F95&partnerID=40&md5=2f0e02da6b0d78aaa6523df4f48308c9

- Zhivkov, P., & Simidchiev, A. (2020). Relationship between Particulate Matter and Health Indicators for Acute Morbidity in Sofia.
  In Proceeding of 1st international Conference on Environmental Protection and Disaster Risks (pp. 180-193).

  https://www.ceeol.com/search/chapter-detail?id=926651

6.2 Декларация за оригиналност

Декларирам, че дисертационният труд съдържа оригинални резултати, получени, при проведени от мен, научни изследвания с подкрепата и съдействието на научния ми ръководител.

Резултатите, които са получени, описани и/или публикувани от други учени, са коректно и подробно цитирани в библиографията.

Настоящият дисертационен труд не е прилаган за придобиване на научна степен в друго висше училище, университет или научен институт.

Подпис:

Библиография

[1] Norfazillah Ab Manan, Azimatun Noor Aizuddin, and Rozita Hod. Effect of air pollution and hospital admission: a systematic review. *Annals of global health*, 84(4):670, 2018.

[2] Stacey E Alexeeff, Noelle S Liao, Xi Liu, Stephen K Van Den Eeden, and Stephen Sidney. Long-term pm2. 5 exposure and risks of ischemic heart disease and stroke events: review and meta-analysis. *Journal of the American Heart Association*, 10(1):e016890, 2021.

[3] Emma Barnes and Marc Schlossberg. Improving cyclist and pedestrian environment while maintaining vehicle throughput: Before-and after-construction analysis. *Transportation research record*, 2393(1):85–94, 2013.

[4] Hannah Bast, Daniel Delling, Andrew Goldberg, Matthias Müller-Hannemann, Thomas Pajor, Peter Sanders, Dorothea Wagner, and Renato F Werneck. Route planning in transportation networks. In *Algorithm engineering*, pages 19–80. Springer, 2016.

[5] Jörg Becker, Daniel Beverungen, Martin Winter, and Sebastian Menne. Umwidmung und weiterverwendung von traktionsbatterien. *Szenarien, Dienstleistungen und Entscheidungsunterstützung*, 2019.

[6] M.L. Bell, J.M. Samet, and F. Dominici. Time-series studies of particulate matter. *Annu. Rev. Public Health*, 25:247–280, 2004.

[7] Alexander Y Bigazzi and Miguel A Figliozzi. Review of urban bicyclists' intake and uptake of traffic-related air pollution. *Transport Reviews*, 34(2):221–245, 2014.

[8] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[9] Iain M Carey, Richard W Atkinson, Andrew J Kent, Tjeerd Van Staa, Derek G Cook, and H Ross Anderson. Mortality associations with long-term exposure to outdoor air pollution in a national english cohort. *American journal of respiratory and critical care medicine*, 187(11):1226–1233, 2013.

[10] N. Castell et al. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment international*, 99:293–302, 2017.

[11] J.W. Cherrie et al. Effectiveness of face masks used to protect beijing residents against particulate air pollution. *Occupational and environmental medicine*, 75(6):446–452, 2018.

[12] Theodoros Chondrogiannis, Panagiotis Bouros, Johann Gamper, and Ulf Leser. Alternative routing: k-shortest paths with limited overlap. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–4, 2015.

[13] Theodoros Chondrogiannis, Panagiotis Bouros, Johann Gamper, and Ulf Leser. Exact and approximate algorithms for finding k-shortest paths with limited overlap. In *20th International Conference on Extending Database Technology: EDBT 2017*, pages 414–425, 2017.

[14] L. Clancy et al. Effect of air-pollution control on death rates in dublin, ireland: an intervention study. *The lancet*, 360(9341):1210–1214, 2002.

[15] Aaron J Cohen, Michael Brauer, Richard Burnett, H Ross Anderson, Joseph Frostad, Kara Estep, Kalpana Balakrishnan, Bert Brunekreef, Lalit Dandona, Rakhi Dandona, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the global burden of diseases study 2015. *The Lancet*, 389(10082):1907–1918, 2017.

[16] Ellen M Considine, Colleen E Reid, Michael R Ogletree, and Timothy Dye. Improving accuracy of air pollution exposure measurements: Statistical correction of a municipal low-cost airborne particulate matter sensor network. *Environmental Pollution*, 268:115833, 2021.

[17] Audrey De Nazelle, Mark J Nieuwenhuijsen, Josep M Antó, Michael Brauer, David Briggs, Charlotte Braun-Fahrlander, Nick Cavill, Ashley R Cooper, Hélène Desqueyroux, Scott Fruin, et al. Improving health through policies that promote active travel: a review of evidence to support integrated health impact assessment. *Environment international*, 37(4):766–777, 2011.

[18] CE Delft. Health costs of air pollution in european cities and the linkage with transport. *CE Delft: Delft, The Netherlands*, 2020.

[19] Inês D Do Vale, Ana S Vasconcelos, and Gonçalo O Duarte. Inhalation of particulate matter in three different routes for the same od pair: A case study with pedestrians in the city of lisbon. *Journal of Transport & Health*, 2(4):474–482, 2015.

[20] F. Dominici, L. Sheppard, and M. Clyde. Health effects of air pollution: a statistical review. *International Statistical Review*, 71(2):243–276, 2003.

[21] Jiyuan Dong, Yurong Liu, and Hairong Bao. Revalue associations of short-term exposure to air pollution with respiratory hospital admissions in lanzhou, china after the control and treatment of current pollution. *International Journal of Hygiene and Environmental Health*, 231:113658, 2021.

[22] Muriel Gevrey, Ioannis Dimopoulos, and Sovan Lek. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological modelling*, 160(3):249–264, 2003.

[23] M. Gochfeld and J. Burger. Disproportionate exposures in environmental justice and other populations: the importance of outliers. *American Journal of Public Health*, 101(S1):S53–S63, 2011.

[24] T. Guan et al. The effects of facemasks on airway inflammation and endothelial dysfunction in healthy young adults: a double-blind, randomized, controlled crossover study. *Particle and fibre toxicology*, 15(1):30, 2018.

[25] A. Gábelová et al. Genotoxicity of environmental air pollution in three european cities: Prague, košice and sofia. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 563(1):49–59, 2004.

[26] David H Hagan, Gabriel Isaacman-VanWertz, Jonathan P Franklin, Lisa MM Wallace, Benjamin D Kocar, Colette L Heald, and Jesse H Kroll. Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments. *Atmospheric Measurement Techniques*, 11(1):315–328, 2018.

[27] Walter Ham, Abhilash Vijayan, Nico Schulte, and Jorn D Herner. Commuter exposure to pm2. 5, bc, and ufp in six common transport microenvironments in sacramento, california. *Atmospheric Environment*, 167:335–345, 2017.

[28] Ghassan B Hamra, Neela Guha, Aaron Cohen, Francine Laden, Ole Raaschou-Nielsen, Jonathan M Samet, Paolo Vineis, Francesco Forastiere, Paulo Saldiva, Takashi Yorifuji, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environmental health perspectives*, 2014.

[29] A.J. Hedley et al. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in hong kong: an intervention study. *The lancet*, 360(9346):1646–1652, 2002.

[30] Mario A Hernández, Omar Ramírez, John A Benavides, and Juan F Franco. Urban cycling and air quality: Characterizing cyclist exposure to particulate-related pollution. *Urban Climate*, 36:100767, 2021.

[31] David M Holstius, A Pillarisetti, KR Smith, and EJAMT Seto. Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in california. *Atmospheric Measurement Techniques*, 7(4):1121–1131, 2014.

[32] Jalil Jaafari, Kazem Naddafi, Masud Yunesian, Ramin Nabizadeh, Mohammad Sadegh Hassanvand, Mansour Shamsipour, Mohammad Ghanbari Ghozikali, Hamid Reza Shamsollahi, Shahrokh Nazmara, and Kamyar Yaghmaeian. The acute effects of short term exposure to particulate matter from natural and anthropogenic sources on inflammation and coagulation markers in healthy young adults. *Science of The Total Environment*, 735:139417, 2020.

[33] Rohan Jayaratne, Xiaoting Liu, Phong Thai, Matthew Dunbabin, and Lidia Morawska. The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. *Atmospheric Measurement Techniques*, 11(8):4883–4890, 2018.

[34] Wan Jiao, Gayle Hagler, Ronald Williams, Robert Sharpe, Ryan Brown, Daniel Garver, Robert Judge, Motria Caudill, Joshua Rickard, Michael Davis, et al. Community air sensor network (cairsense) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern united states. *Atmospheric Measurement Techniques*, 9(11):5281–5292, 2016.

[35] Marilena Kampa and Elias Castanas. Human health effects of air pollution. *Environmental pollution*, 151(2):362–367, 2008.

[36] K. Katsouyanni et al. Short term effects of air pollution on health: a european approach using epidemiologic time series data: the aphea protocol. *Journal of Epidemiology & Community Health*, 50(Suppl 1):S12–S18, 1996.

[37] Christine M Kendrick, Adam Moore, Ashley Haire, Alexander Bigazzi, Miguel Figliozzi, Christopher M Monsere, and Linda George. Impact of bicycle lane characteristics on exposure of bicyclists to traffic-related particulate matter. *Transportation research record*, 2247(1):24–32, 2011.

[38] Kirsten A Koehler and Thomas M Peters. New methods for personal exposure monitoring for airborne particles. *Current environmental health reports*, 2(4):399–411, 2015.

[39] Suk Wah Kwok and Chris Carter. Multiple decision trees. In *Machine Intelligence and Pattern Recognition*, volume 9, pages 327–335. Elsevier, 1990.

[40] O. Laurent et al. Air pollution, asthma attacks, and socioeconomic deprivation: a small-area case-crossover study. *American journal of epidemiology*, 168(1):58–65, 2008.

[41] P. Li et al. The acute effects of fine particles on respiratory mortality and morbidity in beijing, 2004–2009. *Environmental Science and Pollution Research*, 20(9):6433–6444, 2013.

[42] Xiaoting Liu, Rohan Jayaratne, Phong Thai, Tara Kuhn, Isak Zing, Bryce Christensen, Riki Lamont, Matthew Dunbabin, Sicong Zhu, Jian Gao, et al. Low-cost sensors as an alternative for long-term air quality monitoring. *Environmental research*, 185:109438, 2020.

[43] Feng Lu, Dongqun Xu, Yibin Cheng, Shaoxia Dong, Chao Guo, Xue Jiang, and Xiaoying Zheng. Systematic review and meta-analysis of the adverse health effects of ambient pm2. 5 and pm10 pollution in the chinese population. *Environmental research*, 136:196–204, 2015.

[44] Lisha Luo, Yunquan Zhang, Junfeng Jiang, Hanghang Luan, Chuanhua Yu, Peihong Nan, Bin Luo, and Mao You. Short-term effects of ambient air pollution on hospitalization for respiratory disease in taiyuan, china: a time-series analysis. *International journal of environmental research and public health*, 15(10):2160, 2018.

[45] Kamal Jyoti Maji, Wei-Feng Ye, Mohit Arora, and SM Shiva Nagendra. Pm2. 5-related health and economic loss assessment for 338 chinese cities. *Environment international*, 121:392–403, 2018.

[46] Ernesto QV Martins and Marta Pascoal. A new implementation of yen's ranking loopless paths algorithm. *Quarterly Journal of the Belgian, French and Italian Operations Research Societies*, 1(2):121–133, 2003.

[47] L. Morawska et al. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environment international*, 116:286–299, 2018.

[48] Christopher JL Murray. The global burden of disease study at 30 years. *Nature medicine*, 28(10):2019–2026, 2022.

[49] National Statistical Institute of Bulgaria (NSI). Population by cities https://www.nsi.bg/en/content/2981/population-towns-and-sex, 2019.

[50] World Health Organization. *International classification of diseases:[9th] ninth revision, basic tabulation list with alphabetic index.* 1978.

[51] World Health Organization. International classification of diseases (icd) information sheet. 2018.

[52] Luc Int Panis, Bas de Geus, Grégory Vandenbulcke, Hanny Willems, Bart Degraeuwe, Nico Bleux, Vinit Mishra, Isabelle Thomas, and Romain Meeusen. Exposure to particulate matter in traffic: a comparison of cyclists and car passengers. *Atmospheric Environment*, 44(19):2263–2270, 2010.

[53] Hye-Youn Park, Susan Gilbreath, and Edward Barakatt. Respiratory outcomes of ultrafine particulate matter (ufpm) as a surrogate measure of near-roadway exposures among bicyclists. *Environmental Health*, 16(1):1–7, 2017.

[54] S. Pattenden, B. Nikiforov, and B. Armstrong. Mortality and temperature in sofia and london. *Journal of Epidemiology & Community Health*, 57(8):628–633, 2003.

[55] C.A. Pope, D.W. Dockery, and J. Schwartz. Review of epidemiological evidence of health effects of particulate air pollution. *Inhalation toxicology*, 7(1):1–18, 1995.

[56] Xiaoliang Qin, Lujian Hou, Jian Gao, and Shuchun Si. The evaluation and optimization of calibration methods for low-cost particulate matter sensors: Inter-comparison between fixed and mobile methods. *Science of The Total Environment*, 715:136791, 2020.

[57] Aakash C Rai, Prashant Kumar, Francesco Pilla, Andreas N Skouloudis, Silvana Di Sabatino, Carlo Ratti, Ansar Yasar, and David Rickerby. End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Science of The Total Environment*, 607:691–705, 2017.

[58] Sanjay Rajagopalan, Sadeer G Al-Kindi, and Robert D Brook. Air pollution and cardiovascular disease: Jacc state-of-the-art review. *Journal of the American College of Cardiology*, 72(17):2054–2070, 2018.

[59] S Rasoul Safavian and David Landgrebe. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3):660–674, 1991.

[60] J. Samet and D. Krewski. Health effects associated with exposure to ambient air pollution. *Journal of toxicology and environmental health, Part A*, 70:227–242, 2007.

[61] J.M. Samet et al. The national morbidity, mortality, and air pollution study. part ii: morbidity and mortality from air pollution in the united states. *Res Rep Health Eff Inst*, 94(pt 2):5–79, 2000.

[62] David Savage, Xiuzhen Zhang, Xinghuo Yu, Pauline Chou, and Qingmai Wang. Anomaly detection in online social networks. *Social Networks*, 39:62–70, 2014.

[63] J. Shi et al. Cardiovascular benefits of wearing particulate-filtering respirators: a randomized crossover trial. *Environmental health perspectives*, 125(2):175–180, 2017.

[64] S. Shin et al. Ambient air pollution and the risk of atrial fibrillation and stroke: A population-based cohort study. *Environ Health Perspect*, 127(8):87009, 2019.

[65] L.A. Tallon et al. Cognitive impacts of ambient air pollution in the national social health and aging project (nshap) cohort. *Environ Int*, 104:102–109, 2017.

[66] Yan Tao, Shengquan Mi, Shuhong Zhou, Shigong Wang, and Xiaoyun Xie. Air pollution and hospital admissions for respiratory diseases in lanzhou, china. *Environmental pollution*, 185:196–201, 2014.

[67] G.D. Thurston et al. A joint ers/ats policy statement: what constitutes an adverse health effect of air pollution? an analytical framework. *European Respiratory Journal*, 49(1), 2017.

[68] H. Traboulsi et al. Inhaled pollutants: the molecular scene behind respiratory and systemic diseases associated with ultrafine particulate matter. *International Journal of Molecular Sciences*, 18(2):243, 2017.

[69] J. Vermylen et al. Ambient air pollution and acute myocardial infarction. *Journal of Thrombosis and Haemostasis*, 3(9):1955–1961, 2005.

[70] W. Wang et al. Particulate air pollution and ischemic stroke hospitalization: How the associations vary by constituents in shanghai, china. *Sci Total Environ*, 695:133780, 2019.

[71] Scott Weichenthal, Ryan Kulka, Aimee Dubeau, Christina Martin, Daniel Wang, and Robert Dales. Traffic-related air pollution and acute changes in heart rate variability and respiratory function in urban cyclists. *Environmental health perspectives*, 119(10):1373–1378, 2011.

[72] Lingkun Wu, Xiaokui Xiao, Dingxiong Deng, Gao Cong, Andy Diwen Zhu, and Shuigeng Zhou. Shortest path and distance queries on road networks: An experimental evaluation. *arXiv preprint arXiv:1201.6564*, 2012.

[73] Sherrie Xie and Blanca E Himes. Personal environmental monitoring. In *Precision in Pulmonary, Critical Care, and Sleep Medicine*, pages 305–320. Springer, 2020.

[74] Antonella Zanobetti, Francesca Dominici, Yun Wang, and Joel D Schwartz. A national case-crossover analysis of the short-term effect of pm2. 5 on hospitalizations and mortality in subjects with diabetes and neurological disorders. *Environmental Health*, 13(1):1–11, 2014.

[75] R. Zhang et al. Acute effects of particulate air pollution on ischemic stroke and hemorrhagic stroke mortality. *Front Neurol*, 9:827, 2018.

[76] Petar Zhivkov. Optimization and evaluation of calibration for low-cost air quality sensors: Supervised and unsupervised machine learning models. In *2021 16th Conference on Computer Science and Intelligence Systems (FedCSIS)*, pages 255–258. IEEE, 2021.

[77] Petar Zhivkov and Alexander Simidchiev. Development of software tool for optimization and evaluation of cycling routes by characterizing cyclist exposure to air pollution. *Annals of Computer Science and Information Systems*, 32:105–112, 2022.

[78] Petar Zhivkov and Alexander Simidchiev. Software tool for optimizing cycling route by defining cyclist air pollution exposure. In *The Workshop on Computational Optimization*, pages 152–170. Springer, 2024.

[79] N. Zimmerman et al. Closing the gap on lower cost air quality monitoring: Machine learning calibration models to improve low-cost sensor performance. *Atmos. Meas. Tech. Discuss*, pages 1–36, 2017.

[80] Marina Zusman, Cooper S Schumacher, Amanda J Gassett, Elizabeth W Spalt, Elena Austin, Timothy V Larson, Graeme Carvlin, Edmund Seto, Joel D Kaufman, and Lianne Sheppard. Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study. *Environment international*, 134:105329, 2020.

[81] Moniek Zuurbier, Gerard Hoek, Marieke Oldenwening, Virissa Lenters, Kees Meliefste, Peter Van Den Hazel, and Bert Brunekreef. Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route. *Environmental health perspectives*, 118(6):783–789, 2010.

Abstracts of Dissertations

Number 3, 2025

INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES

BULGARIAN ACADEMY OF SCIENCES

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

ИНСТИТУТ ПО ИНФОРМАЦИОННИ И КОМУНИКАЦИОННИ ТЕХНОЛОГИИ

Брой 3, 2025

Автореферати на дисертации