TÉCNICO LISBOA

Air Quality Monitoring and Interpolation System for IoT Applications

Bernardo Nuno Bernardo Amaral

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisors: Prof. João Paulo Baptista de Carvalho
Prof. António Manuel Raminhos Cordeiro Grilo

Examination Committee

Chairperson: Prof. Teresa Maria Sá Ferreira Vazão Vasques
Supervisor: Prof. João Paulo Baptista de Carvalho
Member of the Committee: Prof. João Pedro Castilho Pereira Santos Gomes

December 2019

I declare that this document is an original work of my own authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Dedicated to my family.

Acknowledgments

The first appreciations go to my thesis supervisors, Prof. António Grilo and Prof. João Carvalho, due to all the precious guidance, help and advice provided.

I would also like to thank Eng. Sandra Mesquita for the availability and the important informations supplied, CCDR-LVT stations maintenance employees for the collaboration in the placement of the developed sensors in the monitoring stations. I would also like to extend my thanks to Eng. Luís Lamela from Altice Portugal and Eng. Jose Adarve from Quectel, for the provided debugging help with the new NB-IoT devices, and Altice Golabs.IoT Laboratory for the given opportunity to learn about new IoT technologies in their workshops.

To all my good and remarkable friends Francisco Fialho, Celso Mateus, Raquel Baptista, Filipa Lindo, Maria Belo, Carlos Carvalho, Rodrigo Jacinto, Sara Valente, Carla Vicente, Luís Bucho, Filipa Ribeiro, João Filipe, Magali Correia, and to all that were not mentioned, for their long and lasting friendships and for all the support given during these years.

I would like to express my deep gratitude to NEECIST for all the experiences, friendships and knowledge provided during my college life.

To all my family, most notably to my mother, Adélia Amaral, my father, Pedro Amaral, my brother, Francisco Amaral, my aunt and my godfather, Madalena Bernardo and Manuel Bernardo, and to the rest of my family, for their unconditional daily support and patience.

Finally to my loved one and best friend, Mónica Lourenço, for all her love, care and support.

Abstract

Air pollution is a global problem due to its consequences on population health. World Health Organization estimates that every year 7 million deaths are caused by air pollution related to small inhalable particles. This work intends to provide useful research in the scope of air pollution monitoring and spatial prediction, through the development of a pioneer Narrow-band Internet of Things (NB-IoT) based system for the measurement of particulate matter with a diameter smaller than 10 micrometers, using the low-cost PMS5003 sensor. This system integration in official monitoring clusters was tested, through the comparison with the Portuguese Environment Agency reference sensors in Lisbon, aiming to improve the spatial density of the current network. Furthermore, to enhance the resolution of available data, several algorithms, such as Ordinary Kriging, Inverse Distance Weighting, Linear Interpolation, Nearest Neighbors and Fuzzy Boolean Nets were tested for spatial interpolation through cross-validation, with data from 2013 to 2017. Finally, a web platform for the visualization of live interpolated PM10 concentration was produced. NB-IoT presented low coverage in the city, and comparisons between sensor measurements showed several inconsistencies in the low-cost sensor. Ordinary Kriging and Inverse Distance Weighting, with a zero power function, presented the best performance, indicating a low distance correlation between the reference stations. The developed application allowed a successful representation of high resolution interpolated data. Currently, the NB-IoT technology is not well established in Portugal and further research is needed to improve the integration of low-cost solutions in air quality monitoring networks.

Keywords

Air Quality; Particulate Matter; Spatial Interpolation Models; Fuzzy Boolean Nets; LPWAN; NB-IoT.

Resumo

A poluição do ar é um problema global devido às suas consequências para a saúde da população. A Organização Mundial da Saúde estima que todos os anos sete milhões de mortes são causadas pela poluição do ar. Esta dissertação pretende obter informações cientificamente úteis no âmbito da monitorização e previsão espacial de poluentes ambientais através da construção de um sistema Narrow-band Internet of Things (NB-IoT) para a medida de partículas finas com um diâmetro menor que 10 micrómetros (PM10), utilizando o sensor PMS5003. A integração deste sistema em redes de monitorização oficiais foi testada, através da comparação com sensores de referência da Agência Portuguesa do Ambiente, de modo a melhorar a densidade espacial destas. Para aumentar a resolução dos dados disponíveis, algoritmos como Ordinary Kriging, Inverse Distance Weighting, Linear Interpolation, Nearest Neighbors e Fuzzy Boolean Nets foram testados para interpolação espacial, com dados desde 2013 a 2017. Finalmente, uma plataforma web para a visualização de concentrações de PM10 em tempo real foi desenvolvida. A tencnologia NB-IoT apresentou pouca cobertura em Lisboa, e comparações entre as medidas dos sensores mostraram inconsistências no sensor PMS5003. Ordinary Kriging e Inverse Distance Weight, com função de potência igual a zero, apresentaram o melhor desempenho, indicando pouca correlação entre as estações em termos de distância. A aplicação desenvolvida permitiu a representação de dados com alta definição. Atualmente, a tecnologia NB-IoT ainda não está estabelecida em Portugal e um maior estudo das soluções de baixo preço para a monitorização da qualidade do ar é necessário.

Palavras Chave

Qualidade do Ar; Partículas do ar; Modelos de Interpolação Espacial; Fuzzy Boolean Nets; LPWAN; NB-IoT.

Contents

1 Introduction **1**

1.1 Topic Overview . . . 3

1.2 Motivation . . . 4

1.2.1 Objectives . . . 5

1.3 Claim of Contributions . . . 5

1.4 Contents . . . 6

2 State of the Art **7**

2.1 PM Measurement . . . 9

2.1.1 Gravimetric . . . 9

2.1.2 Optical . . . 9

2.1.3 Microbalance . . . 10

2.1.4 Portuguese Monitoring Station Network . . . 10

2.2 Low-cost Air Monitoring Systems . . . 11

2.2.1 Previous Projects . . . 12

2.2.2 Low-cost PM sensors . . . 13

2.3 Interpolation Algorithms for Spatial Analysis . . . 15

2.3.1 Fuzzy Boolean Networks . . . 15

2.3.2 Inverse Distance Weighting . . . 19

2.3.3 Kriging . . . 19

2.3.4 External Influences . . . 20

2.3.5 Temporal Forecasting . . . 21

2.3.6 Practical Applications . . . 22

2.4 LPWAN Technologies . . . 23

2.4.1 SigFox . . . 24

2.4.2 LoRaWAN . . . 25

2.4.3 LTE-M . . . 26

2.4.4 NB-IoT . . . 26

2.5 IoT Hardware Solutions . . . 27
2.6 Air Pollution Visualization . . . 29
3 PM10 Monitoring, Interpolation and Visualization System **31**
3.1 Low Cost NB-IoT PM10 Monitoring System . . . 33
3.1.1 Development and Composition . . . 33
3.1.2 Placement and Behavior Analysis . . . 36
3.2 Spatial Interpolation Algorithms . . . 37
3.2.1 Dataset Selection . . . 37
3.2.2 Spatial interpolation Algorithms Implementation . . . 39
3.2.3 FBN Implementation . . . 41
3.2.4 Algorithm Performance Evaluation . . . 43
3.3 System for Online Data Visualization . . . 44
3.3.1 Database . . . 44
3.3.2 Mapbox GL JS . . . 45
4 Results **47**
4.1 NB-IoT PM10 Monitoring System . . . 49
4.1.1 NB-IoT Coverage in Lisbon . . . 49
4.1.2 Evaluation of Obtained Measures . . . 50
4.2 SIMs Testing . . . 53
4.2.1 SIM Parametrization . . . 53
4.2.2 SIM Comparison . . . 55
4.3 Online Data Visualization Web Application . . . 57
4.4 Discussion . . . 58
4.4.1 Developed PM10 Monitoring System . . . 58
4.4.2 SIMs Assessment . . . 60
4.4.3 Web Visualization Application . . . 61
5 Conclusion **63**
5.1 Conclusion . . . 65
5.2 Future Work . . . 66

List of Figures

1.1 Air quality monitoring and information state in Portugal. 4
1.2 Development of the IoT PM monitoring and presentation system. 6

2.1 Current monitoring network in the county of Lisbon. 11
2.2 PM concentration measurement methods. 12
2.3 Example of commercialized PM monitors. 13
2.4 Example of the constituents of a FBN. 16
2.5 Example of the learning process of a FBN. 17
2.6 Example of a kriging variogram. 20
2.7 Comparison of LPWAN with other technologies. 25
2.8 Governmental public online platforms for live air quality visualization. 30
2.9 Breezometer high resolution live air quality map. 30

3.1 First prototype of the developed PM monitoring system. 36
3.2 Overview of the developed monitoring system. 36
3.3 Final assembly of the developed monitoring system. 37
3.4 Considered stations in the greater area of Lisbon. 38
3.5 Dataset exclusion process. 40
3.6 Stations in the greater area of Lisbon used in this work. 41
3.7 Spatial grid used for FBN coordinate scale conversion. 42

4.1 Time series of the PM10 concentrations reported by the sensors and corresponding RH. . 51
4.2 Scatter diagram between the developed system and the reference monitor obtained measures. 51
4.3 Measures of the days with the highest and lowest absolute errors and Pearson coefficients. 52
4.4 Scatter diagrams obtained from 43 956 inferences for IDW, OK and FBN, respectively. . . 56
4.5 Histograms obtained from 43 956 inferences for IDW, OK and FBN, respectively. 57
4.6 Developed web application final result, with live data from QualAr, interpolated with OK. . 58

List of Tables

2.1 LPWAN technologies comparison. 28

3.1 AT commands used with the NB-IoT micro-controller. 34
3.2 Bytes in PMS5003 Active Mode functionality. 35
3.3 PM10 data completeness of considered stations before and after data filtering. 39
3.4 Correlation coefficient between target and remaining stations in the Lisbon monitoring network. 39

4.1 Classification of PM10 concentration values. 49
4.2 Weeks during which the developed system was placed next to a CCDR-LVT monitoring container. 50
4.3 Data obtained during the measurement period. 50
4.4 Results of the measurement period, in ($\mu g/m^3$). 50
4.5 Pearson coefficients calculated for the full days of sensor positioning. 52
4.6 IDW power function comparison results. 53
4.7 OK variogram comparison results. 54
4.8 FBN Parametrization assessment results. 54
4.9 Choice of parameters for FBNs used in this work. 55
4.10 FBN 15 minute improvements. 55
4.11 Model comparison results. 56
4.12 Access times of various air quality visualization web pages. 58

Acronyms

3GPP Third Generation Partnership Project

ALF Alfragide

ALV Alverca

ANFIS Adaptive Neuro Fuzzy Inference System

ANN Artificial Neural Networks

APA Portuguese Environment Agency

AVL Avenida da Liberdade

CCDR-LVT Regional Commission for Coordination and Development of Lisbon and Tagus Valley

CPU Central Processing Unit

EEA European Environment Agency

eMTC Extended Machine Type Communication

ENC Entrecampos

FBN Fuzzy Boolean Nets

GIS Geographic Information System

GSM Global System for Mobile Communications

GPRS General Packet Radio Service

GPS Global Positioning System

IDW Inverse Distance Weighting

IoT Internet of Things

LOC Loures Centro

LoRa Long Range

LoRaWAN Long Range Wide Area Network

LPWAN Low Power Wide Area Network

LTE Long Term Evolution

LTE-M Long Term Evolution for Machines

LTE-MC Long Term Evolution for Machine Communications

MAC Medium Access Control

MAE Mean Absolute Error

MEM Mem Martins

NB-IoT Narrow-band Internet of Things

OFDMA Down-link Orthogonal Frequency-Division Multiple-Access

OK Ordinary Kriging

OLI Olivais

PHY Physical Layer

PM Particulate Matter

PM2.5 particles with an aerodynamic diameter smaller than 2.5 micrometers

PM10 particles with an aerodynamic diameter smaller than 10 micrometers

QMA Quinta do Marquês

QualAr Portuguese Online Database on Air Quality

RH Relative Humidity

RMSE Relative Mean Square Error

SC-FDMA Single-Carrier Frequency Division Multiple-Access

SFF Small Form Factor

SIM Spatial Interpolation Method

UDP User Datagram Protocol

UNB Ultra Narrow-band

WHO World Health Organization

1

Introduction

Contents

1.1 Topic Overview 3
1.2 Motivation 4
1.3 Claim of Contributions 5
1.4 Contents 6

In recent years, urban air pollution has been a growing worry among countries and their citizens mainly due to its negative impact on population health and life quality. It is a global challenge for governments, which make big investments to decrease air pollution, to raise the awareness of their citizens and to implement effective policies and interventions to tackle the problem. By monitoring and forecasting air quality it is possible to have a better understanding on its evolution and on how to proceed to improve it [1].

1.1 Topic Overview

Currently, air pollution is a major environmental problem, as it is estimated by the World Health Organization (WHO) that 9 out of 10 people breathe air containing high levels of pollutants and that every year there are 7 million deaths caused by outdoor and indoor air pollution. Outdoor air pollution caused approximately 4.2 million deaths in 2016 and is currently estimated to be responsible for 25% of adult deaths from strokes, 43% from chronic obstructive pulmonary disease and 29% from lung cancer [1]. WHO states that it affects children neural and motor development and damages their lung function, even at lower levels of exposure. This is a worrying situation since globally, 93% of the world's children under 15 years are exposed to particles with an aerodynamic diameter smaller than 2.5 micrometers (PM2.5) above the WHO guidelines [2].

Air pollution also has a negative impact on society in other ways, such as by stunting plant growth, lowering agricultural productivity and by reducing city progress and attractiveness to citizens, slowing their evolution and development [3].

European authorities have approved several directives and council decisions to tackle this problem. In case of non-compliance with these, management measures should be implemented in the affected areas. These plans usually resolve around the use of numerical models that allow the prediction of harmful air quality situations and that should provide data to tackle them. However, since air pollution dispersion is usually complex, air quality models used for modeling and forecasting, in order to be reliable, result in coarse grid forecasts, which in most cases are not useful for the management of critical situations [4].

Systems for the measurement of air pollution currently being used in cities result of the investments made in the last decades addressing the emerging knowledge about air pollution health hazards. These conventional systems are container-style monitoring stations, which occupy large spaces and have high power consumption, requiring high maintenance and production costs. Moreover, its use can only provide a coarse overview of the city's overall air pollution, since the number of sensors per unit area, even in metropolises, is very low, due to the high investment needed for their construction. In the greater Lisbon area, the city with the highest populational density in Portugal, there are only 14 air pollution monitoring

stations, which measure the presence of pollutants such as Ozone and particles with an aerodynamic diameter smaller than 10 micrometers (PM10). From these, only 4 stations measure the presence of PM2.5. In the overall area of Portugal the monitoring station's network completed 72 stations in 2005 [5]. In Figure 1.1 it is possible to see an example of a monitoring station being used in official monitoring networks, as well as how coarse the information attained from this network is in Portugal. The map interface where the data from this network is displayed online is publicly available at Portuguese Online Database on Air Quality (QualAr) website [6].

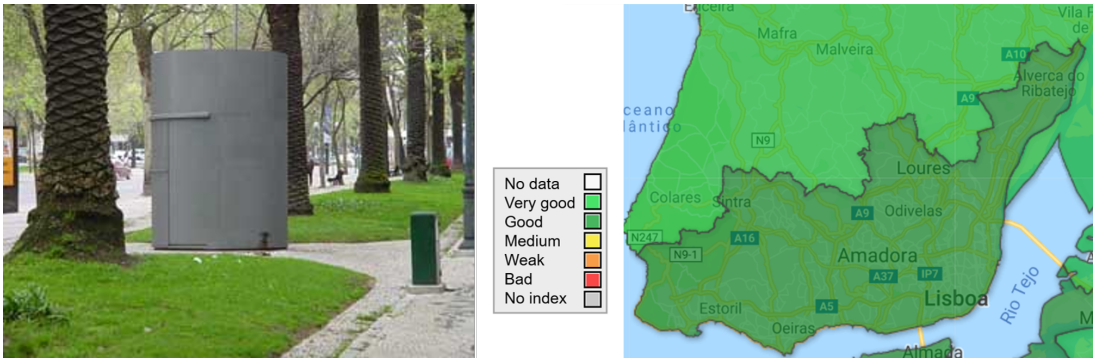

Figure 1.1: Air quality monitoring and information state in Portugal.

1.2 Motivation

Throughout this and the previous decade, technology has seen big developments in the electronic and telecommunication areas, which has made possible the creation of small and low-cost sensors, compared to the ones in the aforementioned conventional systems. These sensors can be combined with electronics and technologies such as mobile communications and the Internet of Things (IoT) to provide cheaper and more effective solutions to the current systems. With the objective of tackling the low amount of resources to model air pollution with a more granular space-time resolution, some governments are already developing initiatives with the use of these new technologies [3].

Modelling and air quality data processing also play a big role in the control of air pollution, by providing tools for a better visualization of its evolution. This knowledge facilitates urban management, through the attainment of risk maps, which are very useful in critical situations. These provide information that raises the awareness of citizens, to change their lifestyle habits, according to the pollution in their cities, and therefore minimizing their own exposure to air pollution [3].

In this work, from all air polluting substances, higher importance will be given to Particulate Matter (PM), which comprises small, solid particles that are a complex mixture of chemical species, originated by a variety of sources, from anthropogenic, mainly combustion processes, to natural, such as storms, pollen and forest fires. According to European Environment Agency (EEA), in 2017, many European

cities have shown concentration values of these particles above the limit values [7]. These particles can penetrate airways, lungs and blood vessels and are known to be responsible for cardiovascular and respiratory diseases as well as lung cancers. Furthermore, PM2.5 particles include pollutants such as sulfate, nitrates and black carbon, which are carcinogenic and pose the greatest risks to human health [1].

1.2.1 Objectives

The goal of this work is to develop a system for an high resolution display of PM10 concentration levels throughout the greater area of Lisbon, with the use of the latest IoT technologies available, spatial interpolation algorithms, machine learning and a web application.

The development of this system will be composed of multiple phases. First, a low-cost portable IoT PM sensing system will be developed. Second, with the help of the Lisbon air quality measures and QualAr data, several Spatial Interpolation Method (SIM)s will be assessed and compared with Fuzzy Boolean Nets (FBN)s in what regards spatial interpolation performance. Finally, a web application which is able to present live, fine resolution, interpolated data on PM10 concentration in the city of Lisbon will be built.

The interpolation algorithms will be evaluated based on the measures taken by the network of monitoring stations in the greater area of Lisbon and which data is available in QualAr. This database comprises validated measurement data from the year 1995 to 2017, depending on the availability of each station of the network at that time.

The development of the aforementioned system aims to tackle the current presentation of data regarding air quality, which at the moment is characterized by its coarse and poor resolution, as can be observed in Figure 1.1.

In Figure 1.2 is presented a flowchart with all the steps that constitute the development of the IoT PM monitoring and presentation system.

1.3 Claim of Contributions

This thesis has the following contributions:

- The development of an usable, affordable, and portable Narrow-band Internet of Things (NB-IoT) system used for measurement of PM10 and PM2.5. In the future, this could be particularly useful for the integration of these affordable systems in the official air pollution monitoring networks.
- An evaluation of FBN interpolation capabilities, as well as a review of state-of-the-art SIMs, in the scope of air quality spatial interpolation applied with the state of information, in the city of Lisbon,

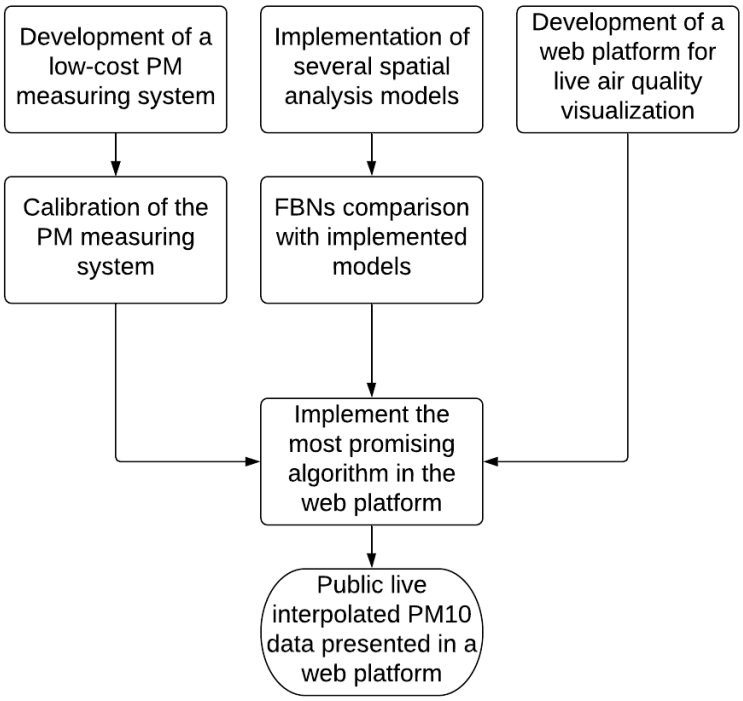

Figure 1.2: Development of the IoT PM monitoring and presentation system.

available at the time of this thesis.

- The development of a web application that is able to present live interpolated PM pollution data in the city of Lisbon, with live data provided by QualAr.

1.4 Contents

Chapter 1 is the Introduction. It contains an overview over the topic of this work, with air pollution being revealed as a global environmental concern. A description of the motivation, objectives and contributions of this work is provided.

In Chapter 2, technologies and studies related to the field of work are reviewed, mainly in what regards performance of low-cost sensors, IoT technologies and SIMs. There is also an overview on the current forms of public presentation of available air quality data.

Chapter 3 presents every step and methodology that will be used in this work, in detail, from the circuit development and algorithm testing, to the building of the web application prototype.

In Chapter 4, the research and development process is concluded with an overview of the results of the built systems and the tested algorithms, along with its discussion and assessment.

In Chapter 5, the main conclusions attained by this thesis are provided, along with the evaluation of its contributions and the future work that can be developed.

2

State of the Art

Contents

2.1 PM Measurement 9
2.2 Low-cost Air Monitoring Systems 11
2.3 Interpolation Algorithms for Spatial Analysis 15
2.4 LPWAN Technologies 23
2.5 IoT Hardware Solutions 27
2.6 Air Pollution Visualization 29

In this section, a state-of-the-art analysis will be made to the current literature and technologies regarding the main subjects of this thesis. These will encompass the existing projects and sensors for the measurement of air quality, specifically regarding PM, the Low Power Wide Area Network (LPWAN) IoT technologies and hardware platforms used along with these in remote systems development, the current SIMs and algorithms used for environmental phenomena estimation and visualization, and finally, the current available solutions at public disposal for the visualization of live air quality, in the form of web applications.

2.1 PM Measurement

There are many types of instruments and methods for the measurement of the concentration and size distribution of PM. These rely on the various behavior characteristics of particles (mobility, aerodynamics and diffusion) [8]. The scope of this thesis only encompasses the measurement of the concentration of PM10 and PM2.5, due to its harm to human health, reviewed in the previous chapter, therefore no size distribution methods will be described.

The main methods for the measurement of PM concentration in ambient air are gravimetric, optical and microbalance [8].

2.1.1 Gravimetric

The gravimetric method is the European reference method for sampling and determining PM10 and PM2.5 in ambient air, as in the norm EN 12341:2014.

In the gravimetric method, particle mass concentration is determined through the usage of filters to weigh particles, before and after the sampling period. It is considered as the basic method to measure PM mass concentrations in combustion gases. The filter collects PM in all granulometric fractions and sizes and then a cyclone or impactor is used to remove larger particles. These filters can only output measures from 15 to 15 minutes, therefore this method is not ideal for the identification of real time events. However, the collected particles can be analyzed chemically, which is not possible in some of the other methods. The filters used in this method are very condition and calibration dependant [8].

2.1.2 Optical

Optical instruments can be based on the principles of light scattering, absorption or extinction, with scattering being the most usual. They usually include dispersion photometers for the measurement of the intensity of scattered light in one or more angles from a combination of all the particles present in the optical detection volume. In the light scattering methods, particles are hit by a light beam and irradiate

that light in every direction. Most commercial light scattering instruments use visible light and measuring angles of 90°, 45°, or less than 30° [8].

2.1.3 Microbalance

In the microbalance method, particles are collected into the surface of an oscillatory microbalance, then through oscillation, these use the alteration of resonance frequency to determine PM concentration. For this method there are two main prevalent instruments, the Tapered Element Oscillation Microbalance and the Quartz Crystal Microbalance.

In the first one, PM mass is measured based on the alteration of resonance frequency of a tapered quartz wand, according to the accumulation of particles in a sampling filter, connected to the wand. In the second, the quartz crystal has a piezoelectric property of changing its resonance frequency when there is a small addition of mass in its surface, with particles being deposited by electrostatic precipitation in a fine quartz crystal resonator, instead of a sampling filter [8].

Amaral et al., in 2015, reviewed every type of PM instrument available in the market, with the goal of helping researchers choose the most suitable equipment for their application. It was concluded that when deciding which device is most appropriate, issues such as detection limit and size range of the instrument must be evaluated [8].

Regarding specific instruments, it was concluded that methods which use filters are not very accurate, but have the advantage of performing chemical analysis in the measured particles. It also requires a lot of labor for it to function properly. Finally, it is also concluded that there was not an absolute best instrument for measuring PM, but a most suitable for each situation, and that different methods should simultaneously be used, in order to confirm values during sampling.

2.1.4 Portuguese Monitoring Station Network

In Figure 2.1 is presented the current monitoring network in the county of Lisbon presented in QualAr website, where it is possible to see the available stations at a certain time, along with color graded measures from the available stations [9]. In Portugal, according to the Manual on Methods and Operational Procedures of Air Quality Monitoring Networks, made by the Portuguese Environmental Reference Laboratory, the methods used for the measurement of PM10 and PM2.5 are the gravimetric and the beta ray absorption [10].

The gravimetric method works as stated in Section 2.1.1. It allows determining the mass of particles deposited in the filters, and these are later used in laboratory to determine particle chemical properties [10].

The monitor used at the stations at the time of this thesis, which was used as reference in this work,

was the Environnement S.A. Model MP101M, which is certified equivalent to the reference method in accordance with the EN12341 European standards for PM10 and PM2.5 particulate concentration. It is based on the beta ray attenuation measurement technique, which is an optical light absorption PM concentration measurement method [11].

A schematic summary of every PM concentration measurement method analysed in this section is presented in Figure 2.2.

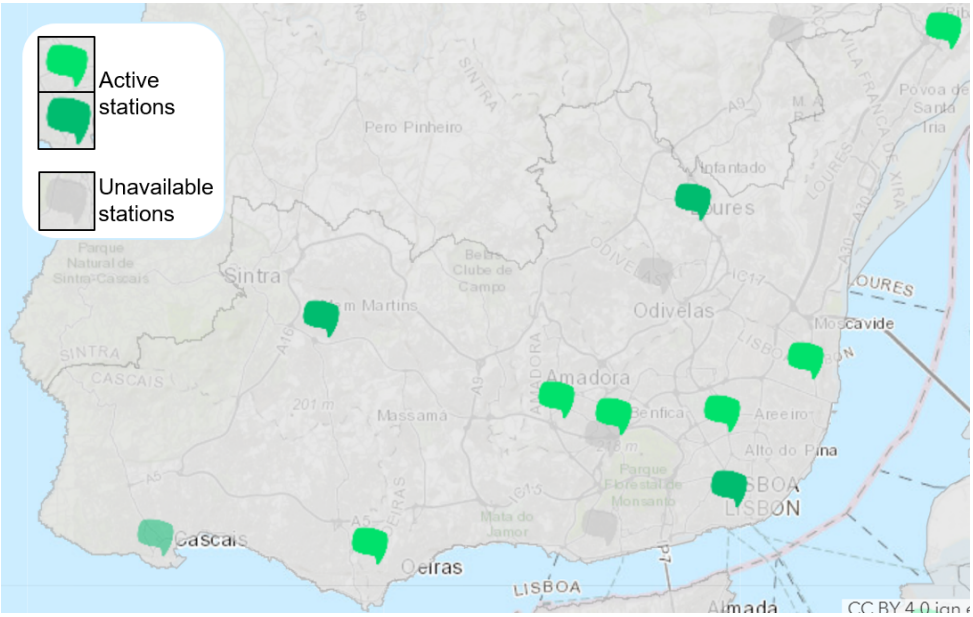

Figure 2.1: Current monitoring network in the county of Lisbon.

2.2 Low-cost Air Monitoring Systems

The scope of this thesis is aimed at the usage of low-cost sensors to further extend current monitoring networks with more data. For this to work, some examples of usage of low-cost sensors were observed in the literature in order to conclude whether these low-cost PM sensors are reliable, taking into account its performance, cost, and calibration needs.

Several models for monitoring systems have been developed with the objective of providing a better solution for the monitoring of air pollution with much lower maintenance and production costs than the current monitoring stations. Many of these use IoT devices and take advantage from mobile communication technologies [12] [13] [14].

In this section, first there will be an overview over research projects and dissertations prior to this thesis where air quality was attempted to be measured at low-costs. Second, it will be shown literature regarding the performance of low-cost PM sensors currently available in the market.

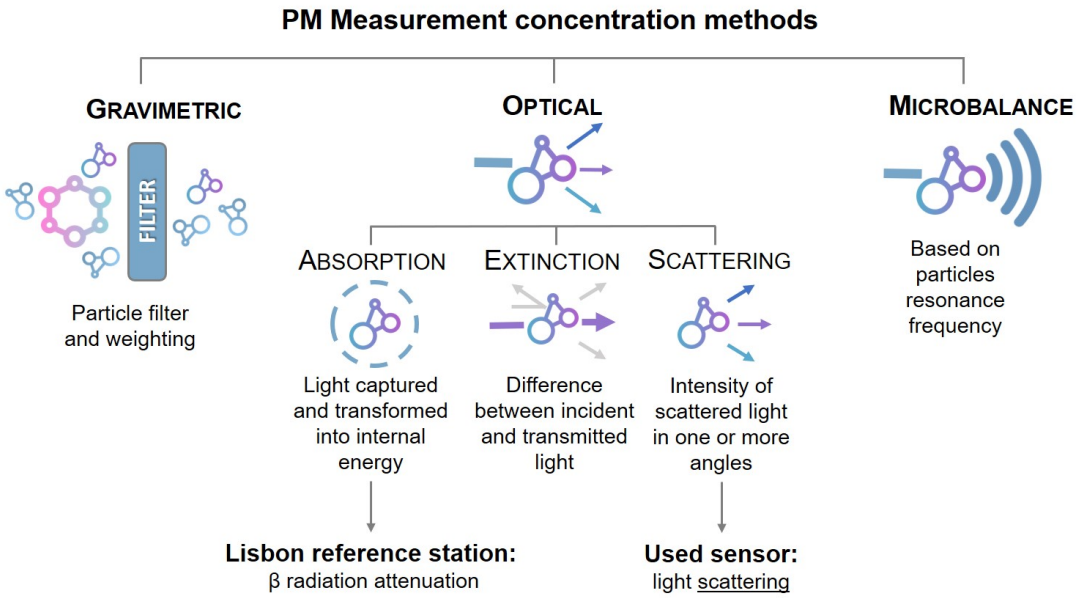

Figure 2.2: PM concentration measurement methods.

2.2.1 Previous Projects

In 2008, Carvalho developed, in his thesis, a mobile system to monitor air quality which could be connected to mobile objects such as a bus or a taxi, in order to acquire a map of the quality of air [15]. This system had a Global Positioning System (GPS) module to provide the location of the system and a Global System for Mobile Communications (GSM) module to transmit data to a central server. The measured gases were CO, NO2, O3, SO2 and CO2, therefore PM measures were absent. Each module was self-sufficient and supplied by solar panel charged batteries, and the gases measures were presented in the server in a numerical or geographical form overlapped with a map of the city of Lisbon. However not many campaigns were done, with not much data being collected, and the system's sensors were not calibrated with the reference standard measuring stations.

In 2011, in the context of an improvement of air quality project named URBISNET, which took place in Lisbon, made by several researchers from Instituto Superior Técnico, a prototype of a mobile station for air quality mapping sensor network was developed [14]. This prototype also measured many polluting gases, used GPS, significant processing power, data storage, many inputs and outputs for the different sensors, and GSM. Despite the efforts on the development of this solution, there were problems with the calibration and overall validation of some of the low-cost gas sensors, which rendered the prototype not very useful and with relatively high production costs for the functions provided.

Vehicle and Propulsion Systems Laboratory, currently operating in Instituto Superior Técnico, in its proposal of creation, with the coordination of Tiago Farias, proposed a project for monitoring pedestrians exposure to PM [16]. This consisted of a 11 kg bag, with a GPS module, a micro-controller, a partic-

ulate and dust sensor, a numerical pad and a computer. The collection of particle concentration data combined with nearly instant measures and the live GPS monitoring allow the estimation of the pollution pedestrians are exposed to throughout certain routes.

2.2.2 Low-cost PM sensors

Low-cost solutions for PM measurement currently in the market consist mostly of optical instruments. These are made of small lasers placed in an certain angle, against a photodiode, which is used to detect the scattered light of illuminated particles, as stated in Section 2.1.2. The generated light scattering signal is filtered and amplified. Afterwards a modulation signal is used to represent the measurements. These sensors usually cost less than 50 euros, which is a considerable difference in relation to reference instruments, which have much more expensive production and maintenance costs.

Despite their price, low-cost optical PM monitors can be used to characterize PM concentrations with high spatial and temporal resolution [17]. In 2016, Manikonda et al. conducted a study with the objective of assessing these monitors and concluded that they perform with adequate precision for indoor air monitoring, although with some calibration requirements [17]. Currently, many professional small factor air quality monitoring stations, for both indoor and outdoor applications, are available in the market. Additionally, communities have been formed with the objective of creating networks where individuals can collect and share data from their personal monitoring systems so that all the obtained data is presented collectively in a map. An example of this is the Weather Underground community [18].

Some of these monitoring stations are known to use low-cost and compact size light-scattering based PM sensors, namely, the stations manufactured by Gaia Earth Sensing Labs, which use specifically Plantower PMS sensors [19]. These manufacturers have many product variants for outdoor monitoring. One of these products is presented in Figure 2.3.

Figure 2.3: Example of commercialized PM monitors.

A study from the University of Utah, in 2018, performed a long-term field evaluation of the Plantower PMS sensors, in which for 320 days, multiple season variations, and significant PM2.5 variation events, measurements were undertaken by two sensors, PMS5003 and two PMS1003, with tracking made through co-located reference air monitors. The results of this study demonstrated overall good correlations between the PMS sensors and reference monitors in the winter season, seasonal differences in sensor performance, some intra-sensor variability, and drift in one sensor, with worse results associated to the PMS1003 sensor [20].

Kuula et al., in 2019, investigated how sensors can be used in urban areas to complement existing air quality networks, through a campaign during the winter, with the use of the optical sensors PPD60PV and PPD42NS. These were also compared to reference instruments [21]. The results showed that these sensors could measure PM10 and PM2.5 with reasonably high correlations. It was concluded that comprehensive understanding of the sensor characteristics is necessary, and that the sensor must be validated and calibrated in the environment where it will be placed. It is also stated that one big disadvantage of these sensors is the non-guaranteed accuracy and the lack of established scientific literature. Results also show that optical sensors are suitable for measuring mass concentrations of particles that are larger than 0.5 μm, and that diffusion charging based sensors, instruments which use light extinction to monitor particle concentrations, are better for traffic exhaust and local combustion related ultra fine particles.

Despite being promising, the performance of these low-cost PM sensors under field conditions is still not well understood. In order to characterize the performance capabilities of a low-cost PM sensor, PMS3003 from Plantower, Zheng et al., in 2018, performed a study where these sensors were tested in both low and high particle concentration regions, during different seasons, to determine how the sensors perform across a range of concentrations and meteorological factors [12]. The attained results showed that Relative Humidity (RH) has a major influence in the sensor measures, while the temperature effects were negligible. With that, calibration correction factors were applied to measures taken by the sensor, and the mean errors were reduced from 27% to 10% at different time resolutions, when compared to high precision reference instruments. After calibration, the results proved that the sensor can be a promising addition to current PM sensor networks.

The South Coast Air Quality Management District, agency responsible for South California air quality monitoring, in 2018, field-tested Plantower PMS5003 PM sensor, and notable correlation and precision was attained between the PMS5003 and the reference instruments [22].

2.3 Interpolation Algorithms for Spatial Analysis

Air quality analysis relies in mathematical approaches to explain the polluting gases concentration evolution. The process of developing this type of prevision with high spatial resolution is a rather complex problem, due to the wide sources of pollution in urban and industrial areas, and to all the variables that can influence air quality.

Approaches to this problem in the literature are sometimes deterministic, which require a lot of knowledge on the local sources and factors that can influence air pollution, or otherwise rely on measurement data. Spatial interpolation consists in estimating the value of measures of certain properties, at unsampled locations, within the area covered by existing observations of the same properties. The objective is to present a much more detailed overview of the air quality in a certain region.

Currently, SIMs are well developed and frequently used in various Geographic Information System (GIS) applications, in which they can be used to manipulate, analyze and present data in spatial representations. Spatial interpolation is based on the fact that observation points that are closer to each other have more similarities than the ones far away, which is known as Tobler's First Law of Geography [23].

Some of the most popular and standardized SIMs, since the beginning of spatial interpolation modeling, are Inverse Distance Weighting (IDW), kriging, shape functions, spline, and trend surface. These occur from different approaches to this problem, such as local neighborhood, geostatistical and variational [24].

In 2014, Li and Heap made a review of the 25 most commonly applied SIMs [25]. As result an easy to use decision tree was created for use case dependent method selection. According to it, the most suitable interpolation algorithms in circumstances such as the ones in this work are IDW and Ordinary Kriging (OK), depending on data selection and availability.

2.3.1 Fuzzy Boolean Networks

FBNs are boolean networks developed and studied by researchers at Instituto Superior Técnico [26]. In order to correctly explain the functioning of these, a proper nomenclature needs to be given to its constituents.

The networks are constituted by neuronal areas. Each area has a set of neurons, and the number of neurons, which is user defined, must be equal in all areas. Each neuron has a binary value and is constituted by a binary table with a user defined length. The value at each index of each table is considered a memory of the respective neuron, and all memories are initialized with a random number between zero and one. The number of memories of each consequent neuron is equal to a granularity factor.

Each neuronal area can only be one of two types, antecedent or consequent, with their corresponding neurons named after the area, antecedent or consequent neurons respectively. FBNs must have at least one neuronal area of each of the two types. In these networks, connections are weightless, temporary, circumstance dependent and only appear between neurons located in different areas of different types, either in the learning or in the inference process. In Figure 2.4 a very simple example of an FBN is presented, with two antecedent areas and one consequent area, each with four neurons. The granularity factor is two and consequently, neurons in the consequent area have four memories each. In this representation dark neurons have the value one, and white neurons have the value zero.

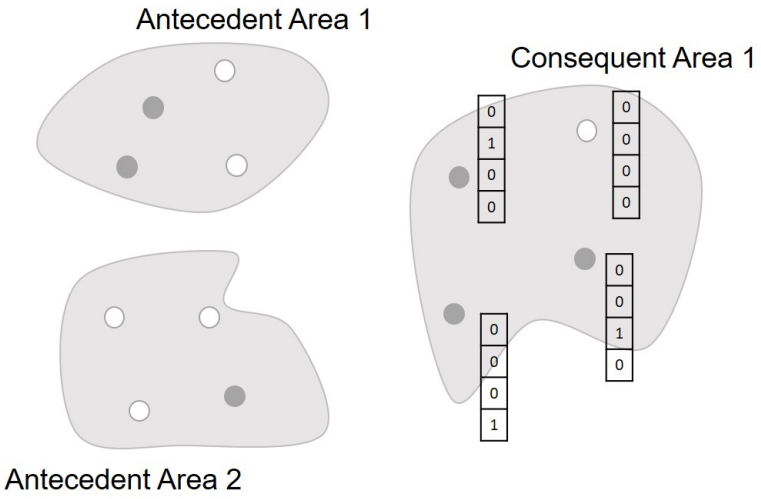

Figure 2.4: Example of the constituents of a FBN.

In what regards the usage of these networks, the number of areas is defined by the number of variables in the problem. To each independent variable, an antecedent area is created and associated. The same happens with the dependent variables, for which consequent areas are created and respectively associated. As an example of usage, if used for linear interpolation, the FBN would only have an antecedent area, associated to the x axis (independent), and a consequent area, associated to the y axis (dependent). The network represented in Figure 2.4 has two independent and one dependent areas, being the latter the subject of inference.

Before the creation of the network, the definition of its parameters must be made. These parameters are the size of each neuronal area, the number of memories per neuron, the granularity factor and the number of samples each neuron uses in the inferring and learning processes. Each should be previously tested and assessed comparatively by the researcher, to each different use case [27].

The learning process of these networks is composed of several phases:

- First, for each network there will be sets of values of each variable in the problem. These are considered the rules that the network will learn. For each set, and for each variable, its value is

translated into a percentage. Each corresponding area will be activated according to this percentage, that is, the neuron values for that area will be 1 with the probability of that percentage, and rest will remain at zero.

- Second, each neuron, in the consequent areas, samples an user defined number of neurons from each antecedent area. This number must always be the same for every neuron and antecedent area, and counts the number of one valued neurons for each area.

- For each count that neuron attained, it multiplies that count by the granularity factor raised to the power of the number of the antecedent area of the respective count.

- Finally, each consequent neuron stores its own binary value (which was defined during the consequent area initial activation) in its memory table, in the index corresponding to the previously acquired count.

Various iterations of this process, for different training values for every variable, represent how training is performed in FBNs. Assuming each memory was initialized with the value zero, it is possible that the represented FBN has already been trained with exactly one set of values, since one memory per activated consequent neuron has the value of one. In this single learning process step, one could conjecture that the connections for that step were as portrayed in Figure 2.5, according to the activated indexes of the table of memories of each consequent neuron (connections with the zero valued neuron in the consequent area were not placed in the figure for ease of visualization, since they are not relevant).

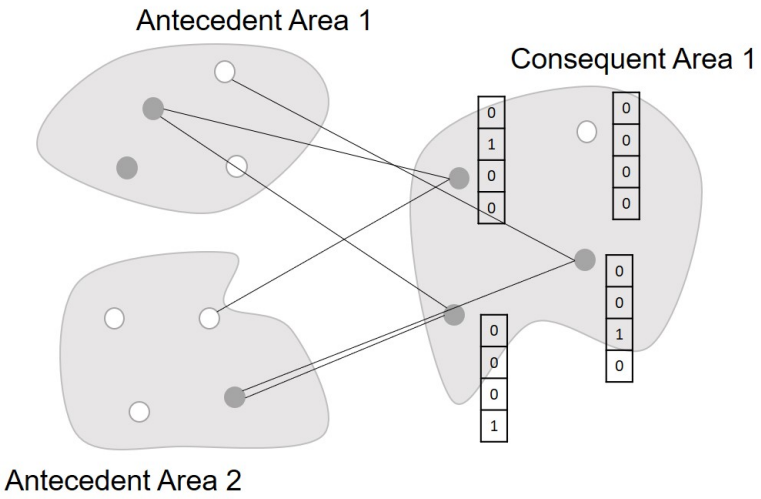

Figure 2.5: Example of the learning process of a FBN.

In the inference process, the network only receives the value of the independent variables and it is supposed to infer the dependent variables value:

- Initially, only the antecedent areas are activated according to its variables value translated into percentage.

- Afterwards, in the exact same fashion of the learning process, every consequent neuron performs sampling, ending up with a count for each antecedent area. The value in the corresponding memory of that neuron is verified. If it stores a zero, then that neuron value will be modified to zero, if it stores a one, its value will be modified to one.

- In the end of this process every neuron in every consequent area will have a binary value. Therefore, consequent areas can be translated into a percentage according to its activation ratio.

- This percentage is translated into the scale of the corresponding dependent variable, and it represents the inference value for the respective variable, in the respective set, according to the respective learning rules.

Studies have concluded that these networks have many nature and biologic like characteristics, such as the organization of neurons per area, to which a given variable or concept can be associated. They have random individual connections and structured meshes of links between them. They are also characterized by relative noise immunity and the capability to approximate reasoning and learning. Provided the remaining neurons are enough to define activation ratios accurately, any number of neurons or connections may be deleted or corrupted [27].

It has also been shown that the FBN learning process converges for a repetitive sequence of different rules, where in the limit, the system is able to learn a number of rules compatible with a granularity that increases with the square root of the number of inputs per neuron and per antecedent [28].

As Universal Approximators, FBN can learn multi-input single-output functions, and they perform pretty well in comparison to other techniques when parametrization is rather difficult. FBN perform better when the training data is very sparse, and data-sets are imbalanced [29]. They are shown to be immune to individual neuron or connection errors, which does not happen with Artificial Neural Networks (ANN), and also have good generalization capabilities [29]. However, one problem that arises from the use of FBN is the that the memory necessity increases exponentially with the number of inputs and size of the network, which is a notable limiting factor.

Tomé et al., in 2014, in a deep analysis of FBN, despite assessing its main characteristics and advantages, also made experiments in which learning, and interpolation capabilities were compared, by interpolating sparse data. The comparisons were made between FBN, Adaptive Neuro Fuzzy Inference System (ANFIS), ANN, Cubic spline interpolation, Support Vector Machines and Isotonic Regression Model, and it was shown that FBN outperformed all other interpolation algorithms. Even when used as a classifier, FBN outperformed every other algorithm, given a sparse set of data [27].

2.3.2 Inverse Distance Weighting

IDW is a deterministic SIM based on the assumption that the degree of influence of the nearby sample points should be greater than the effect of more distant points. Its maximum and minimum value can not occur outside the sampled points. The only parameter which is user defined in the IDW implementation is the power function, which defines the relevance of the distance to the weights of each observed value [30].

The IDW interpolation can be defined by the equation (2.1), where (2.2) represents the weights associated to each observation.

$$f(x,y) = \frac{\sum_{i=1}^{n} w(d_i) z_i}{\sum_{i=1}^{n} w(d_i)}, i = 1, 2, ..., n \tag{2.1}$$

$$w(d_i) = \frac{1}{{d_i}^p} \tag{2.2}$$

Where z_i is the observed value, d_i is the distance between the estimation and the observation points, $w(d_i)$ represents the weight associated to observation *i*, and *p* is the power function.

In IDW, weights are proportional to the inverse of the distance elevated to the power of p. Therefore, as distance increases, weights decrease, in a rate that increases the higher the value of p. When p is zero, the resulting estimate is simply equal to the spatial average of all sampled points.

Even though IDW is one of the most widely used interpolation models, it has several limitations, such as its inability to produce error statistics, unlike OK, and the fact that spatial relationships between two locations are not only dependent on distance neither are they constant over space [30].

2.3.3 Kriging

Kriging interpolation models are fundamental geostatistic tools in the field of spatial analysis.

The main difference between kriging and other non-geostatistical models is the assumption that the spatial correlation structure of the process under study is known and can be estimated from the observed data. In kriging, weights are based not only in the distance between points but also in the overall spatial arrangement of observation points [30].

The kriging schemes are stochastic, local, gradual and exact interpolators. Its methodology includes two stages: the analysis of the spatial variation and the estimation of the target variable, which is also based on the weighted average approach. Furthermore, it uses several statistic models which allow the construction of a variety of maps, including previsions, errors of previsions and probabilities, which are considerable advantages in comparison to IDW.

Before kriging is used, an exploratory statistical analysis of data must be made, as well as the

modelling of a variogram to represent how semivariance varies with distance. The analysis of the spatial variation is performed through the variogram modelling of how distance variations with semivariance are not linear, but have a custom relation. In Figure 2.6, an example of a variogram is presented. The nugget represents the semivariance value when the distance tends to zero. The sill is the point where semivariance stabilizes, and range is the interval in which as distance increases, higher is the semivariance. Kriging interpolation weights are chosen using the modeled variogram, so that estimates are unbiased and the estimation variance is minimized.

The most common kriging models are simple kriging, which assumes a known constant mean, OK, which assumes that there is an unknown constant mean, estimated from the data and universal kriging, which assumes that there is a trend in the surface that partly explains the data variations [31].

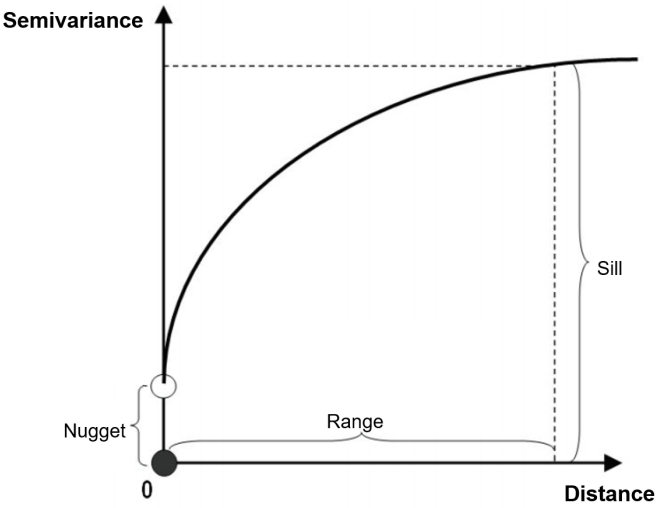

Figure 2.6: Example of a kriging variogram.

2.3.4 External Influences

AERMOD and ADMS-urban are air pollution dispersion models that are reference standards for governments such as the ones of United States of America and United Kingdom for the precision modelling of air quality [32]. AERMOD was developed at the United States Environmental Protection Agency and ADMS-Urban was developed by Cambridge Environmental Research Consultants. Both are used for regulatory purposes in the respective countries and highly recommended models in many other.

AERMOD is a gaussian plume air dispersion model which heavily takes meteorological and other external factors into account. The fact that it is plume based model, means that it considers the movement of one fluid through another, which in this case are the particles which move through the air. Some of the external factors which it takes into account are cloud cover observations, wind speed and direction,

temperature, dew point, humidity and sea level pressure. Meteorological data is accepted from multiple heights and wind, temperature and turbulence are treated as vertical profiles.

ADMS-Urban is a model of dispersion of pollutants released from industrial, domestic and road traffic sources in urban areas. This local Gaussian type model is nested within a trajectory model for areas within 50 km $\times$ 50 km [33].

These models recognize that air turbulence and dispersion are continuously variable, depending on height above ground, winds and temperature. Concerning the height above ground, virtually all airflow near the ground is neutral. Both models have steady state by nature, which means polluting gases conditions, between the time of emission and the time it reaches any receptor, do not change. Therefore, it is clear that plume models cannot be expected to give reasonable results for long range transport of pollutants. Despite this, they can still be used to estimate pollutants concentration in areas where there is no monitoring stations, in the likes of interpolation methods [33].

In 2009, Mesquita, studied the performance of Multiple Linear Regression as an interpolation algorithm that uses external factors like city traffic emissions in comparison to the most common interpolation algorithms which do not encompass external factors, and this approach reflected notably better results than algorithms such as IDW and OK [30].

2.3.5 Temporal Forecasting

Zhang et al., in 2012, made a complete two part review of the history, techniques, current status of science and research and future prospects of real time air quality forecasting, where three major techniques for forecasting are distinguished [34] [35].

The simple empirical approaches, which are based on the assumption that today's observed pollutant level is tomorrow's forecasted value, are considered weak and of low accuracy. There are also parametric and statistical methods, which encompass regression methods and ANNs, and provide moderate to high accuracy. Finally, there are the advanced, physics based approaches, which are computationally expensive and complex and have very high operational costs and also have moderate to high accuracy [34].

It is important to notice that for real time air quality forecasting, most methods are very circumstance dependent and regional related, with a big knowledge needed on the area subject to the forecasts.

Real time air quality forecasting does not belong to the scope of air spatial interpolation, therefore it will not be subject of study in this thesis but it could be considered for a prospect of future work for the developed systems.

2.3.6 Practical Applications

The above mentioned approaches have been developed and experimented both theoretically and empirically, with many of its applications documented in recent literature in the air pollution field.

A review of SIMs for environmental scientists made in 2008, by Li and Heap, theoretically compares 26 SIMs and concludes that the selection of an appropriate method for each different case is critical, but difficult, since its performance depends on the variable under study, the spatial configuration of the data and on the underlying assumptions [24].

In 2013, Funenga, in his masters dissertation, used the OK interpolation method to infer the air quality related to polluting gases in the city of Lisbon, with the data available from the fixed monitoring stations [36]. The results showed poorly defined spatial continuity, since the interpolation only accepted inputs from 6 different stations, which represent very sparse data. Funenga also tested the Calpuff model for interpolation, which was, until 2017, the preferred system for the simulation of air pollution dispersion of the United States Environmental Protection Agency. This consisted of a model for the meteorological factors influence, a Gaussian puff dispersion model and a post-processing program for results output. In comparison to the OK, this model relative mean absolute error was lowered, which demonstrates high performance increase from the Calpuff method usage.

Tang et al., in 2017, proposed a spatial interpolation framework to incorporate diverse data sources and model the spatial processes explicitly at multiple resolutions with the use of spectral analysis and a spatial Gaussian process for interpolation, underlining the fact that when monitoring data sources are very sparse, auxiliary data should be added to the interpolation algorithms [37].

In 2018, Oteros et al., based on weather information, and on a kriging geospatial interpolation method, estimated daily pollen concentration in unmonitored areas, based on 26 monitoring stations in an area of approximately 71 000 km^2. It concluded, that the developed system was accurate enough for the spatial interpolation of pollen concentrations at unmonitored sites [38]. Also in 2018, a study was conducted with the objective of proposing a low-cost mobile sensor network for evaluation of air quality in urban areas, with the use of kriging interpolation for data processing. It concluded that better results could have been taken if the modelling relied not only on measurements but also on spatial information such as weather data, building heights, and population density. It also confirmed that sparsely distributed sensor nodes, along with kriging interpolation, can be used to produce a real-time rendering air quality map [39].

Machine learning techniques have also been incorporated into air quality interpolation, in order to improve its results and some of its processing and data constraints.

Mishra et al., in 2015, used an artificial intelligence approach to forecast PM2.5 during haze episodes, compared Neuro-Fuzzy techniques with ANNs and Multiple Linear Regression, and concluded that its Neuro-Fuzzy model outperformed all the other approaches in the forecasting of PM2.5 [40].

In 2018, Alimissis et al., evaluated both ANN and Multiple Linear Regression as interpolation methodologies, using data from a real urban air quality monitoring network of the concentration of different air pollutants in Athens, Greece. It was concluded that ANN were significantly better, especially where the monitoring network density was limited and there was consequently a low degree of correlation between the monitoring sites [41]. Also in 2018, Franceschi et al. developed models to forecast particulate matter concentration using ANN, which were proven to be useful as references to issue early warnings of high air pollution in certain areas [42].

Measurement sensors are typically sparsely located, and that is one of the biggest problems to SIMs for the prediction of air pollution.

2.4 LPWAN Technologies

Small devices with Central Processing Unit (CPU), memory, sensors and capable of mobile communications through which they can connect to the internet have a great potential to change our society. The inter-networking of these devices is called IoT. Data collected can be shared and analyzed in real time through the Internet. These objects and networks can help improve society and its citizens lives in many different ways.

In fact, the number of devices connected to the Internet has seen a big growth in the recent years and is expected to keep growing exponentially, as people increase the devices they purchase. This growth is unprecedented in the communication industry as well as in the wider global economy. It is estimated that by 2020, more than 50 billion devices will be connected through radio communications [43].

IoT has seen various applications in many practical research fields, due to specific requirements such as long range, low data rate, low energy consumption and cost effectiveness.

The most widely used radio technologies nowadays are either short-range technologies such as Zigbee, Bluetooth and Wi-Fi (which are not suited for scenarios that require long-range transmission), or cellular communication technologies, such as 2G, 3G and 4G (which provide long-range access but have excessive power consumption) [44]. With this gap in the radio communications and the emerging IoT technologies, increasing number of devices and many useful practical application projections, a new group of technologies was developed so that long-range transmission could be provided with low consumption and communication costs, named LPWAN [44]. This group of technologies is considered to be ideal for the research to be developed in this work, with its application to the system that will be produced. It can be used since a low data rate is allowed, and long-range access together with low-consumption are needed.

LPWAN represents the current trend in the evolution of IoT technologies as it is increasingly gaining popularity both in industrial and research environments, due to its low production and consumption

costs. In order to achieve these specifications, usually a small data rate limitation is imposed. LPWAN technologies are an excellent choice for long-term monitoring applications in urban areas, as can be seen by its advantages in Figure 2.7.

Until the appearance of LPWAN technologies, the two main approaches to IoT were multi-hop mesh networks with short-range technologies in the unlicensed spectrum or long-range conventional cellular technologies such as GSM and General Packet Radio Service (GPRS) [45]. Experimental trials made by Zanella et al. in 2016 allowed the conclusion that by using LPWAN technologies, less than half of the gateways are needed to cover the same urban area than the number of sites deployed by one of the major cellular operators in Italy, to provide mobile cellular access.

Currently, there are several LPWAN technologies being used both for research and business applications. The most popular and suitable for this work are SigFox, Long Range Wide Area Network (LoRaWAN), Long Term Evolution for Machines (LTE-M) and NB-IoT. While SigFox and LoRaWAN are proprietary technologies, operating in the unlicensed spectrum, LTE-M and NB-IoT are standardized by Third Generation Partnership Project (3GPP) and operate in the licensed spectrum. In this section, an overview of these technologies is made, as well as a comparative analysis regarding related empiric studies in the literature and the characteristics that differentiate them from each other.

The business and commercialization application of LPWAN is already taking place. A survey on LPWAN technologies, made in 2017, stated that the successful implementation of LPWAN technologies such as NB-IoT and LoRaWAN had already been started in North America and Europe. Meanwhile, several mobile operators in countries like Korea, Japan and China, according to the survey, also already had plans to implement nationwide NB-IoT or LoRaWAN [46].

In Portugal, the mobile operator Altice Portugal has announced total NB-IoT coverage in Portugal by the beginning of 2019 [47].

The rest of the LPWAN technologies have the following coverage in Portugal, at the time of this thesis: SigFox has full coverage, according to its website [48]; LoRaWAN has decent coverage, available through a total of 43 gateways implemented throughout the country by several communities in Portugal [49]; there is currently no record of any LTE-M implementation.

2.4.1 SigFox

The SigFox physical layer employs Ultra Narrow-band (UNB) wireless modulation, and its network layer protocols are proprietary. Bidirectional communication is supported and SigFox claims that each gateway can handle up to a million connected objects, with a coverage area of 30 to 50 km in rural areas and 3 to 10 km in urban areas.

The ultra-narrow-band technology has communication channels with approximate bandwidth of 100Hz, uplink-triggered transmissions that follow a client-server model and cooperative reception from multiple

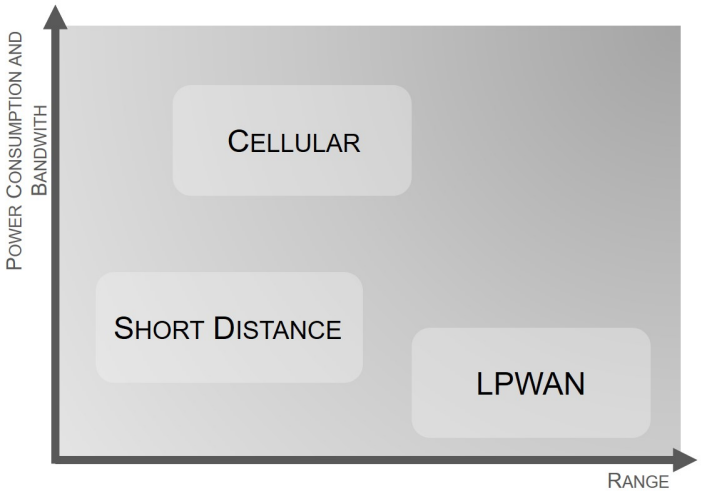

Figure 2.7: Comparison of LPWAN with other technologies.

gateways. An end device wakes up only when uplink data arrives, and then randomly selects a channel for uplink transmission, which is repeated three times to increase the chances of a successful reception by the network. Following this uplink transmission, for the next 10s, there is the possibility for receiving a downlink message on a predetermined channel [50].

2.4.2 LoRaWAN

LoRaWAN is another popular LPWAN solution, patented by Semtech Corporation. It has a proprietary Physical Layer (PHY), Long Range (LoRa), but the rest of the protocol stack, designated LoRaWAN, is open. LoRaWAN networks, like SigFox, typically have a star topology with many end devices connected by a single-hop link to gateways, which are connected to a common network server via standard IP protocols, with its gateways being totally transparent to end devices, providing simpler network management.

LoRaWAN supports 3 types of end devices, each type characterized as a class. Class A supports bidirectional communications through, right after uplink transmission, two short period downlink receive windows. This class has the lowest power consumption, but like SigFox, for downlink transmission, servers need to wait for the next scheduled uplink transmission. Class B devices are also bi-directional, but with scheduled receive slots, and Class C devices have maximum receive slots, which offer the lowest latency for server to device communications in all the classes, but the most power consumption as well.

In order to maximize a device battery and overall capacity, the communication bandwidth is managed by the LoRaWAN network, for each device, which does not happen with SigFox, where the bandwidth is fixed. The LoRa modulation is proprietary and is based on chirp spread spectrum to increase the

immunity to interference in the unlicensed bandwidth [51].

2.4.3 LTE-M

LTE-M, also known as Long Term Evolution (LTE) Cat M1, Long Term Evolution for Machine Communications (LTE-MC) or Extended Machine Type Communication (eMTC), is a standardized LPWAN technology, by 3GPP, which operates in the licensed spectrum. It has the same design of LTE, regarding transmission structure, downlink Down-link Orthogonal Frequency-Division Multiple-Access (OFDMA), uplink Single-Carrier Frequency Division Multiple-Access (SC-FDMA) and same channel coding as well. In order to reduce the cost and power consumption in comparison to LTE, technologies like reduced peak rate, simplified hardware and narrow band operation were implemented [50].

As such, LTE-M has the LTE minimum system bandwidth of 1.08 MHz and the minimum communication bandwidth of 180 kHz per device. This is considered to be wasteful in terms of spectral efficiency and also prevents the deployment in narrower cellular bands like GSM legacy cellular services bands, which have increasingly been re-farmed for new wireless services. It also raises the problem of lack of support to the upcoming industry massive IoT, considering the scarce spectrum [50].

2.4.4 NB-IoT

NB-IoT, also known as LTE Cat NB1, is also a LPWAN technology standardized by the 3GPP, created in 2015 to face the growing massive IoT connectivity challenge, targeting the exploitation of the re-farmed GSM spectrum. It operates on 180 kHz bandwidth (same as LTE bandwidth per device) for both downlink and uplink. The downlink remains with the same structure as LTE (OFDMA with 15 kHz subcarrier spacing), and the uplink is SC-FDMA with sub-carrier spacing at 3.75 kHz (15 kHz for in-band to avoid interference from other LTE traffic), which is also the minimum transmission bandwidth for a device [52].

The advantages of NB-IoT are its enhanced indoor coverage, the ability to connect a massive number of low-throughput devices, but with a lower data rate. Its main objectives include low-cost devices, high coverage, long device battery life, and massive capacity, with some relaxation in latency.

The NB-IoT network is designed to work seamlessly on existing GSM and LTE networks within the licensed frequency bands, making use of the already deployed mobile operator's base stations. Due to this, currently many telecommunication manufacturers are supporting the standardization of the NB-IoT network [53].

A literature analysis regarding empiric and review studies which contain comparisons between the IoT technologies was made. In 2016, Centenaro et al. conducted a LoRaWAN deployment test in Northern Italy, from which it was concluded that the LPWAN paradigm has the potential to complement

current IoT standards as an enabler of smart city applications, since with LoRaWAN, the gateways needed to cover a certain area are considerably less than the base stations a cellular operator would need to cover the same area. A study by Yang et al. in 2017, which analyzed and compared the effective bandwidth of SigFox, LoRa, LTE-M and NB-IoT, concluded that it is yet too early to say which is the best technology, since there still remain many challenging issues waiting for solutions in the area. A Medium Access Control (MAC) layer-based evaluation with the use of a probabilistic model, concluded that NB-IoT is more robust that both LoRaWAN and SigFox in terms of Packet Error Rate [54].

In a study by Ikpehai et al. in 2018, through the specifications analysis and a network simulation scenario of several LPWAN technologies, it was concluded that LoRaWAN offers the longest range, LTE-M has the higher power consumption, and NB-IoT provides the widest coverage, since it is supported by mobile operators, while LoRaWAN gateways are private and sparsely scattered. A study by Lauridsen et al. in 2017, through a coverage comparison of GPRS, NB-IoT, LoRaWAN, and SigFox in a 7800 km area, using a 3GPP Rural Macro and Urban Macro non-line-of-sight model simulation in a realistic scenario, also concluded that from the LPWAN technologies NB-IoT is the one that provides the best coverage, even in indoor scenarios [55].

Other interesting conclusions also were that both SigFox and LoRaWAN are better when the amount of data that must be sent daily is beneath 1000 bytes, while NB-IoT is more suitable for cases where this number is between 1000 and 10000 bytes, since it maintains its extended device lifetime in these situations [56]. A survey specifically on LoRaWAN and NB-IoT technologies concluded that LoRaWAN is more suitable for industrial applications, while NB-IoT is better to personal and public applications [45].

The analysis to the specifications of all the LPWAN technologies mentioned, as well as conclusions attained from the reviewed literature, are presented in Table 2.1, which is inspired in the listed characteristics by Yang et al., in 2017 [50], and summarizes the analysis made in this section.

2.5 IoT Hardware Solutions

In March of 2018, Rahman et al. developed and implemented an IoT based platform for developing cities aimed at environmental monitoring. For this project, 3G cellular connections were used, due to the lack of availability of other low-cost radio technologies. This system was composed of an Arduino Nano micro-controller to retrieve the data from the gas sensors, and a Raspberry Pi to store and process that data, which is then stored and sent through 3G to the cloud [58].

In fact, throughout the literature, many IoT solutions produced for research purposes use single board computers such as Raspberry Pis as the core data processing unit, as well as micro-controllers such as Arduinos. These devices are popular research tools in the development of IoT systems in areas such as healthcare, smart home automation, smart agriculture and environment monitoring [58].

Table 2.1: LPWAN technologies comparison.

	SigFox	LoRaWAN	LTE-M	NB-IoT
Receiver Sensitivity	-147 dBm	-147 dBm	-132 dBm	-137 dBm
Frequency Band	Sub-GHz ISM	Sub-GHz ISM	Licensed	Licensed
Minimum Transmission Bandwidth	100 Hz	125 kHz	125 kHz	3.75 kHz
Modulation	D-BPSK	LoRa proprietary modulation, GFSK	BPSK, QPSK, 16QAM	$\pi \div 2$–BPSK, $\pi \div 4$–QPSK
MAC	Unslotted ALOHA	Unslotted ALOHA	SC-FDMA	SC-FDMA
Over the air updates	No	Yes	Yes	Yes
Related Work	Suitable when data to be sent daily is beneath 1000 bytes [56].	Suitable to IoT industrial applications [46] and when data to be sent daily is beneath 1000 bytes [56]; Longest range [55]	Most heavy power consumption [55].	Suitable to IoT personal and public applications [46] and when data to be sent daily is between 1000 and 10,000 [56]; Widest coverage [55] [57]; The most robust, considering the Packet Error Rate [54].

In recent literature, there are examples of the use of Raspberry Pis as core systems in air quality monitoring research. Ibrahim et al. developed, in 2015, an IoT Smart Monitoring system using a Raspberry Pi, to process and send the data received to the cloud. In 2016, both Shete et al. and Balasubramaniyan et al. developed a similar system to monitor climate data also with the use of a Raspberry Pi [59] [60]. These projects have the limitations of having a relatively high power consumption and the need to use Wi-fi for the Internet access. In 2017, Patil et al. developed a system for the smart monitoring of the air quality also based on a Raspberry Pi. This project was developed using the wireless communication protocol Zigbee and an Arduino micro-controller [61].

Both Raspberry Pis and Arduino based devices are reliable products to be used in IoT systems. Raspberry Pi is a single board computer, running an operating system. An Arduino based micro-

controller is simpler, with much lower memory and processing power, and can only run one program at a time. This results in a higher energy consumption and cost in the Raspberry Pi, and the opposite in comparison to Arduino based micro-controllers. However, in terms of processing power and data storage, the Raspberry Pi is dominant, since Arduino has less memory and processing power [60].

It can be concluded that Raspberry Pis are suitable for heavy duty situations such as, in the air quality monitoring systems scope, the processing and storage of a large amount of data from many sensors. On the other hand, an Arduino based micro-controller would be suitable for a simpler implementation with only one or two sensors and a simple data retrieval and processing mechanism.

Currently, NB-IoT technologies are in early stages of availability, but there are already several solutions provided by the industry for its hardware implementations in the form of Arduino based boards. Main manufacturers of NB-IoT chipsets and modules include Huawei, Intel, QualComm, Samsung, Nordic Semiconductor, Sierra Wireless and U-Blox, and many of these solutions are already being implemented in business domain applications [62].

Despite this, a relevant application of this technology to air quality monitoring systems could still not be found in the literature, providing this work an opportunity to enlighten some information on this area and provide useful conclusions regarding this technology implementation in a low cost and consumption air quality monitoring system.

2.6 Air Pollution Visualization

Air pollution visualization is one of the objectives of this thesis. Research was made in order to find the current forms of public presentation of live information regarding the air quality in parts of the world where there are enough monitoring stations for it to be possible. Currently there are two types of platforms for the visualization of live air quality data: the governmental platforms and the private companies platforms.

Reviewed governmental platforms in Europe and America, do not show fine interpolated data in their air quality maps. They only present, for each individual monitoring station, its local air quality index value. This leads to low resolution maps which are difficult to perceive and do not show air quality variances between stations. Some of these governmental platforms are presented in Figure 2.1. At the top left is the QualAr platform, at the top right is the United States Environment Protection Agency one, at the bottom right is the air visualization provided by EEA, and at the bottom right is the platform provided by the United Kingdom government [9] [63] [64] [65].

Air quality visualization platforms from external companies to the government are business oriented and take advantage of the publicly available air quality data provided governments to make their live interpolated air pollution map as a product. Examples of these are Breezometer and AirVisual. These present heat maps with a much finer resolution than the ones provided by governments, as can be seen

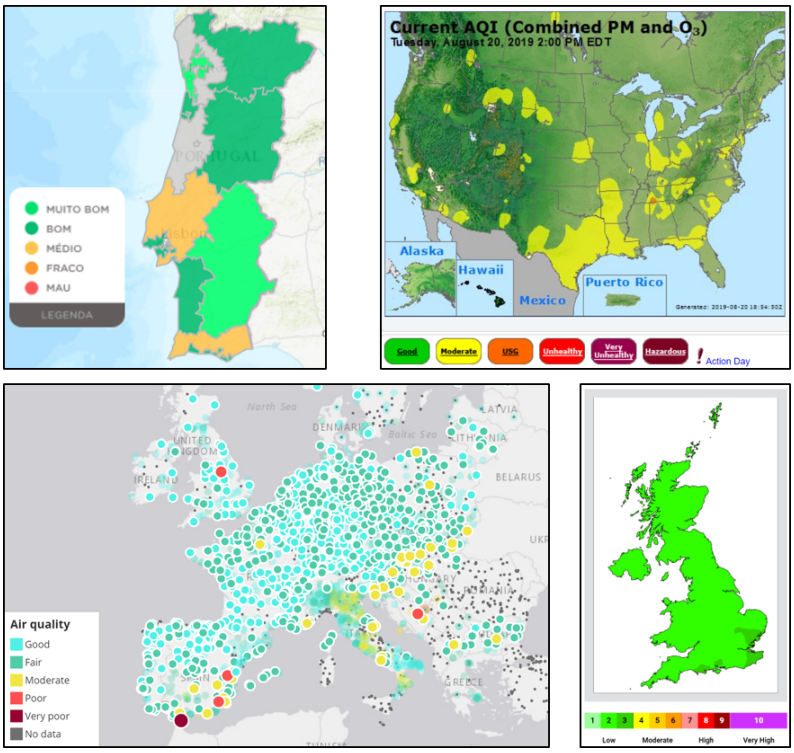

Figure 2.8: Governmental public online platforms for live air quality visualization.

in Figure 2.9 [66]. However, the spatial modelling algorithms used for the interpolation are proprietary, are not open-source and do not contribute to the literature, which is the aim of this work. Additionally, the accuracy provided from these is only extended to 500 meters [67].

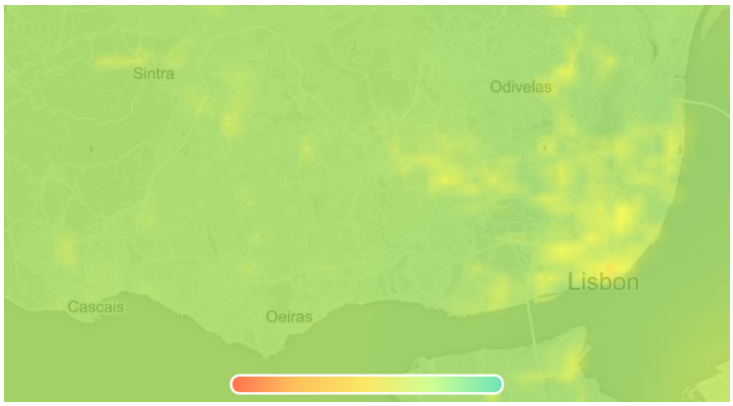

Figure 2.9: Breezometer high resolution live air quality map.

3

PM10 Monitoring, Interpolation and Visualization System

Contents

3.1 Low Cost NB-IoT PM10 Monitoring System 33

3.2 Spatial Interpolation Algorithms 37

3.3 System for Online Data Visualization 44

The methodologies and implementation details concerning the systems developed in this work are presented and explained in this section.

First, a small system that measures PM10 and retrieves them through NB-IoT to a server, which saves them in a database, was developed. This system consists of an NB-IoT module for communication and a PM low-cost sensor. Its performance was compared with one of the official standard PM sensors used by Portuguese Environment Agency (APA) in standard monitoring stations in Lisbon.

Second, SIMs were tested and assessed with the use of data from QualAr, from 2013 to 2017, with a focus on its comparison against FBNs, which have been shown to be promising when interpolating with sparse datasets [27].

Finally, a web application for the presentation of live high-resolution PM pollution rates was built, within a similar area to the city of Lisbon, covering the stations that provide data for the interpolations.

3.1 Low Cost NB-IoT PM10 Monitoring System

The NB-IoT PM10 monitoring system was developed to evaluate how well these type of equipment could integrate current standard air monitoring networks, expanding its size, and consequently providing better data for spatial modelling approaches.

A low cost sensor was used to measure PM10 concentration in the air at a constant rate with relatively low power consumption. Sensor outputs were retrieved and processed by a programmable micro-controller.

The main requirements of this system were low production costs, a Small Form Factor (SFF), low power consumption as well as long range transmission. A low bit rate is acceptable, since the measurement rate is also relatively low. The processing power required is low, since a low amount of data is processed by the micro-controller.

3.1.1 Development and Composition

NB-IoT was the LPWAN techology used in this work, due to the agreements between INESC-ID and Altice Portugal and NOS for its usage for research purposes, and considering the technology characteristics presented in Chapter 2.

SODAQ SARA SFF R412M

The development board used in this work is the SODAQ SARA SFF R412M board, which is an industry standard IoT developer board, with low power consumption, which allows the use of both NB-IoT and LTE-M networks, implying coverage across Europe, North America, Africa and Asia, with 62.5 kbit/s upload and 27.2 kbit/s download rates, and Over-The Air firmware updates [68]. This board contains

an integrated micro-controller, the Atmel SAMD21, which allows it to be programmed through the same tools as every Arduino compatible board.

Subscriber identity module cards provided by the mobile operators NOS and Altice Portugal were used for this board remote communications with NB-IoT. These card were previously configured by the mobile operators according to the specifications needed for its optimal use in NB-IoT devices.

AT commands are used to communicate through the NB-IoT module. The commands used in the implementation of the developed system, and in the micro-controller program, along with its purpose are presented in Table 3.1. These are available in the SODAQ AT command manual [69].

Table 3.1: AT commands used with the NB-IoT micro-controller.

Command	Functionality
AT+CFUN	Sets the mobile terminal to full functionality
AT+CGATT	GPRS Attach
AT+USOCR	Create a socket
AT+USOST	Writes the specified amount of data to the remote address
AT+USOCL	Close socket

The code developed for the micro-controller was written in the Arduino IDE environment, and consists of an internal loop that is continuously receiving measures from the sensor, and every 15 minutes, it makes the averages of the measures taken in that period and sends it through the NB-IoT AT commands, if there is network coverage.

The server code that receives the data from the NB-IoT modules is written in Python. In this program, an User Datagram Protocol (UDP) socket is used to receive packets from the sensor nodes. These contain the sensor location, id, a PM10 and a PM2.5 value, corresponding to the respective 15 minute averages. After receiving the data, it is stored in a database directly from the server.

Plantower PMS5003

A sensor from Plantower was used for the measurement of PM concentration, specifically the PMS5003, a low-cost light-scattering optical sensor with a built-in fan that provides air circulation, widely used and reviewed in the literature [21] [12] [70] [13] [71] [17] [72] [20]. This sensor measures both PM10 and PM2.5 concentration in suspended particles in the air, and outputs it through a digital signal, processed by a microprocessor, in real-time. Its minimum distinguishable particle diameter is 0.3 micrometers. Since it provides a low cost and low consumption profile, with an SFF, and also compatibility with Arduino boards, this sensor is optimal for this use case.

Its power supply requirements are 5V DC, and a maximum of 100mA. It can withstand operating temperatures from -10ºC to 60ºC, as well as relative humidity from 0 to 99%. It is documented by the

manufacturer that its maximum consistency error is of 10 µg/m^3 if the measured values range from 0 to 100 µg/m^3, and of 10% from 100 to 500 µg/m^3.

The PM sensor has two modes of functioning, the active and the passive mode. In the active mode, the sensor is always working and constantly sending digital data to the micro-controller. The data transport protocol works by sending 32 bytes periodically, of which each byte represents a different value. The first four and last four bytes are used for error checking. Bytes from 5 to 10 represent the PM concentration of suspended particles with the maximum diameter of 1, 2.5 and 10 respectively, for factory environments. Bytes from 11 to 16 have the same meaning, but in this is case for atmospheric environments. Finally, remaining bytes represent the number of particles with diameter beyond a certain value, which is a metric that is not in the scope of this thesis. This mode is depicted in Table 3.2. In this work, only the bytes from 11 to 16 are of interest.

Table 3.2: Bytes in PMS5003 Active Mode functionality.

Byte Interval	Functionality
1 to 4	Error check
5 to 10	PM10, PM2.5 and PM1.0 concentration under factory environment
11 to 16	PM10, PM2.5 and PM1.0 concentration under atmospheric environment
16 to 28	Number of particles with diameter beyond 10, 5, 2.5, 1, 0.5 and 0.3 µm in 0.1 liters of air.
29 to 32	Error check

The passive mode is an host transport protocol, in which the micro-controller makes requests to the sensor and it replies accordingly. Due to the requirements on the sensor end, the active mode was the mode used in this work.

Step Up Voltage Converter

For the assembling of the SODAQ SFF R412M board with the PMS5003 sensor, a step up voltage converter is needed, since PMS5003 internal fan requires 5V DC power supply, and the SODAQ board can only output 3.3V. Despite this, the sensor high level of data pins is 3.3V. Therefore, the first iteration of this system can be presented as the one in Figure 3.1.

Similar to the behavior of the official monitoring networks, once data is collected from the sensors, 15 minute averages are calculated and sent to a server, which stores them in a database.

The developed system workflow is depicted in Figure 3.2.

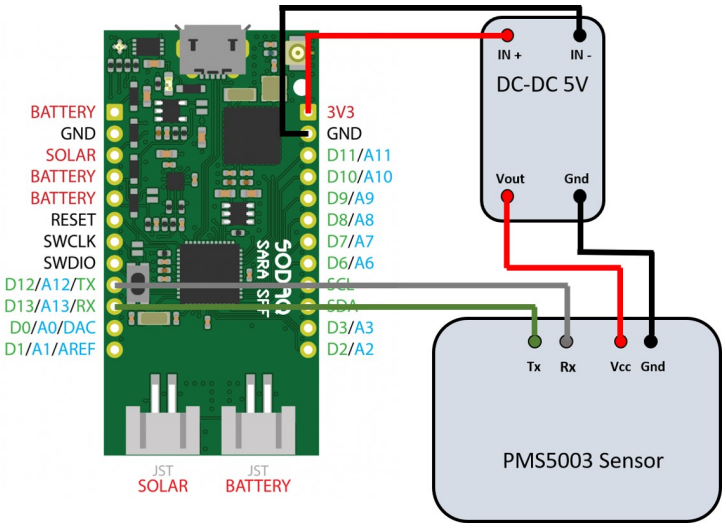

Figure 3.1: First prototype of the developed PM monitoring system.

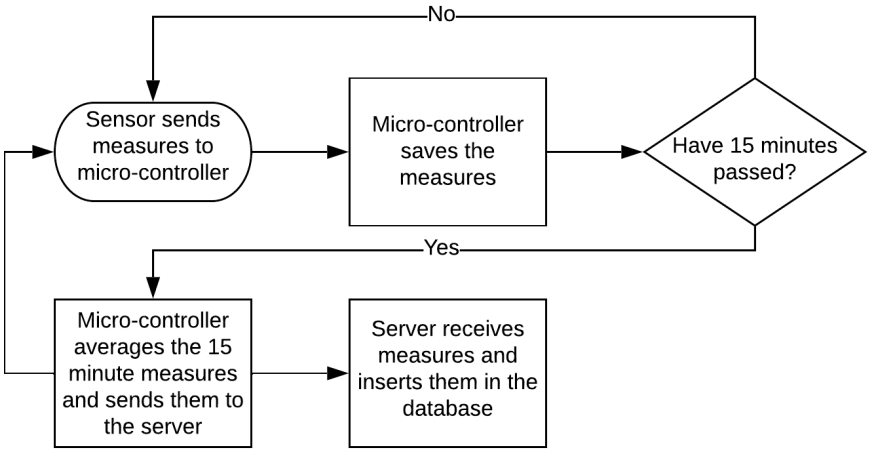

Figure 3.2: Overview of the developed monitoring system.

3.1.2 Placement and Behavior Analysis

Both Entrecampos (ENC) and the Avenida da Liberdade (AVL) monitoring stations were used for the placement of the developed sensors.

Reference PM monitors are placed inside a container, and in order for air to reach the monitors, several teflon pipes are placed outside, within 2 meters of relative altitude, and the air is sucked through these pipes into the monitor container.

Similar behavior was intended to be replicated with the developed system with the junction of a vacuum pump and a plastic pipe to the developed system. However, to have better network coverage, the developed sensor had to be placed outside, in the top of the monitoring stations, next to the air entrance of the station and without any pipe, in a small box resistant to weather conditions and with an

index of protection of 65. In Figure 3.3, the final assembly of this system is presented.

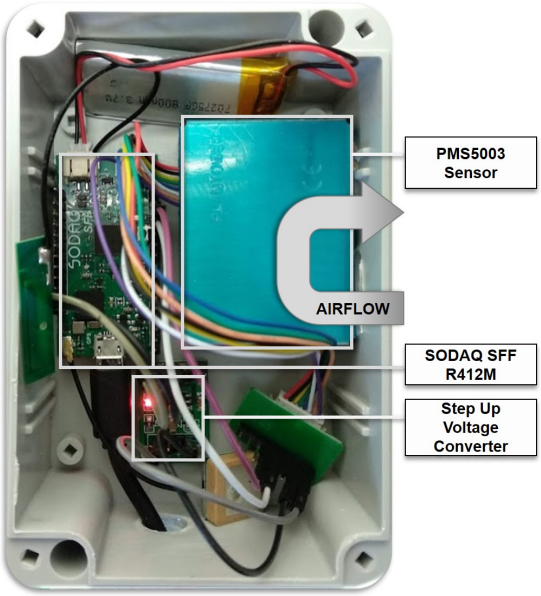

Figure 3.3: Final assembly of the developed monitoring system.

The developed system was placed during two weeks of July in AVL station, four weeks of August and two of September in ENC station and, finally, during the two final weeks of September in AVL station.

3.2 Spatial Interpolation Algorithms

The main algorithms which are assessed in this work are FBN, OK and IDW. Additionally, simpler interpolation algorithms such as LI and NN are also tested. Data from 2005 to 2017 was extracted and analysed from QualAr website for the algorithms assessment. This data only concerns monitoring stations in the greater area of Lisbon,which corresponds to the area where air quality monitoring stations have the most spatial density in Portugal.

In this section, there is an overview of the data selection process, the algorithms parametrization and the tests that were done to assess the algorithms.

3.2.1 Dataset Selection

In the greater area of Lisbon there is an air quality monitoring network composed of 14 monitoring stations, maintained by Regional Commission for Coordination and Development of Lisbon and Tagus Valley (CCDR-LVT) [9]. Only 13 of these have operating PM10 sensors, from which only 10 are fully operating at the time of this work. Only 4 out of these have operating PM2.5 sensors.

PM2.5 data was not regarded due to the lack of stations in the area of analysis.

Only data from 2013 to 2017 was gathered since it is the most consistent overall, with measures from the considered stations, at high percentages of completeness (Table 3.3).

From the five years of measures, a dataset of 43 506 hourly averaged measures was collected for each Lisbon PM10 measuring station. However, there were still many periods with incomplete data from some of the stations, since these were not working during certain periods. In order to remove data inconsistencies, the dataset was filtered based on the number of data per time interval. Each set was required to have at least 10 measures from all the considered stations, and at least 3 measures from the stations in the center of the network, presented in red in Figure 3.4.

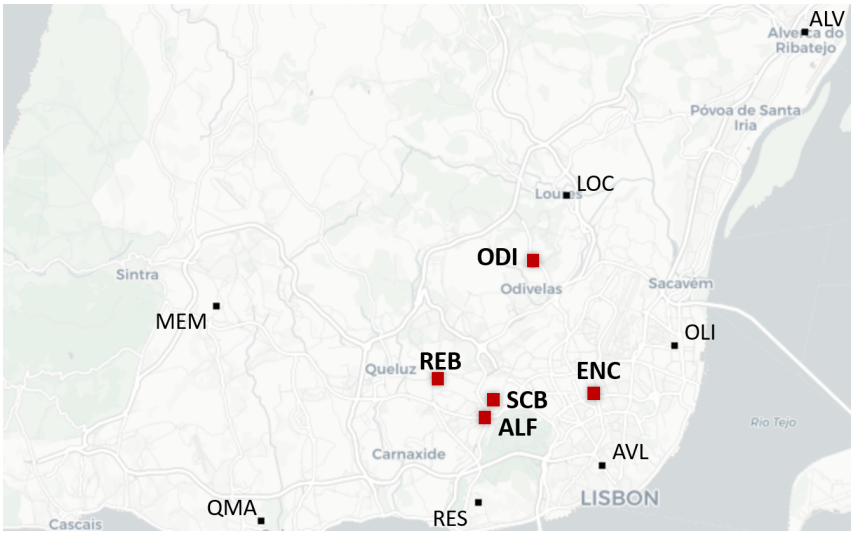

Figure 3.4: Considered stations in the greater area of Lisbon.

In Table 3.3, the percentages of completeness of each station before and after dataset filtering are presented. An increase of data completeness is evident in all stations. The Alfragide (ALF) station demonstrated very low completeness overall, before and after filtering, therefore it was removed from the dataset.

The Alverca (ALV) station was also removed from the dataset, because it is located too far from the center stations, which would increase the problem sparsity and possibly degrade the algorithm performance. This data exclusion process can be observed in Figure 3.5, where *N* represents the size of each dataset.

The number of sets of measures in the dataset decreased from 43 506 measures in the initial dataset, to 11 657 in the filtered one. A map with the remaining Lisbon monitoring stations, after dataset filtering, is presented in Figure 3.6.

Data correlation was also calculated between every station and the center target stations. These coefficients help to explain the quality of the spatial prediction obtained by the various models. In Table 3.4 these are presented, with an highlight of the stations which share higher linear correlation.

Table 3.3: PM10 data completeness of considered stations before and after data filtering.

Station	ID	Station Type	Area Type	Data Completeness (%)	
				Overall	Filtered
Alfragide/Amadora	ALF	Background	Urban	1.40	2.85
Alverca	ALV	Background	Urban	97.47	99.96
Avenida da Liberdade	AVL	Traffic	Urban	96.49	99.21
Entrecampos	ENC	Traffic	Urban	76.40	99.50
Loures-Centro	LOC	Background	Urban	85.88	91.46
Mem Martins	MEM	Background	Urban	88.31	99.72
Odivelas-Ramada	ODI	Traffic	Urban	82.89	98.88
Olivais	OLI	Background	Urban	93.74	99.54
Quinta do Marquês	QMA	Background	Urban	89.30	99.55
Reboleira	REB	Background	Urban	68.88	96.88
Restelo	RES	Background	Urban	29.66	40.81
Santa Cruz de Benfica	SCB	Traffic	Urban	48.67	82.00

Table 3.4: Correlation coefficient between target and remaining stations in the Lisbon monitoring network.

	ENC	ODI	REB	SCB
AVL	0.83	0.77	0.78	0.76
ENC	1	0.84	0.82	**0.81**
LOC	0.77	0.79	0.79	0.69
MEM	0.77	0.78	0.8	0.66
ODI	**0.84**	1	**0.86**	0.78
OLI	0.83	0.76	0.78	0.74
QMA	0.77	0.78	0.83	0.71
REB	0.82	**0.86**	1	0.8
RES	0.76	0.83	0.81	0.73
SCB	0.81	0.78	0.8	1

These were calculated with the Pearson correlation coefficient, which has been used in similar applications [31], and represents the linear correlation between two variables. The possible result values vary from 1 to -1, where 1 represents total positive linear correlation, 0 represents no linear correlation, and -1 total negative linear correlation.

3.2.2 Spatial interpolation Algorithms Implementation

In order to assess algorithm performance, cross validation tests were applied to the selected stations. In each iteration, one measure from one central station (Figure 3.6) was inferred using the measures from remaining stations in the same time interval. In Algorithm 3.1 the pseudo-code of the setup of this process is presented.

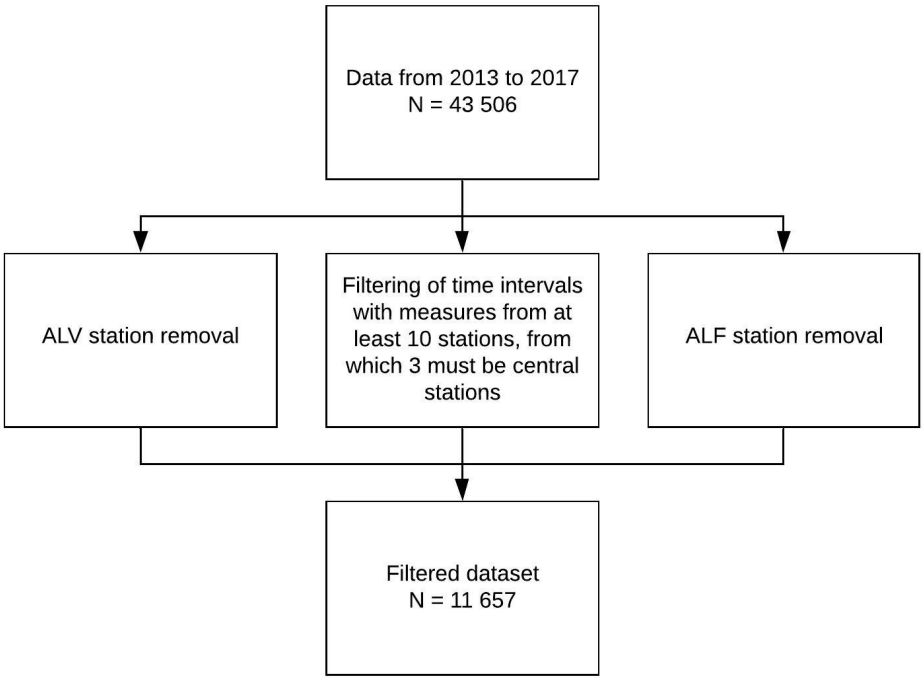

Figure 3.5: Dataset exclusion process.

Algorithm 3.1: Pseudo-code of the algorithm execution setup.

```
dataset = extractDataset();
stations = stationsData();
centralStations = centralStationsData();
foreach set in dataset do
    foreach station in stations do
        if station in centralStations then
            executeInterpolationAlgorithms(station, set);
        end
    end
end
```

Both LI and NN were implemented with the python *scipy* library *interpolate* package.

In what regards IDW, its only user defined parameter is the power function value p, which defines how the weight of each observed point varies with the distance to the estimated point. This algorithm was implemented in Python. Cross validation was applied to decide which is the best value of p for the selected data.

OK is the Kriging variant which was considered in this work, and it was implemented with the *pykrige* python library [73].

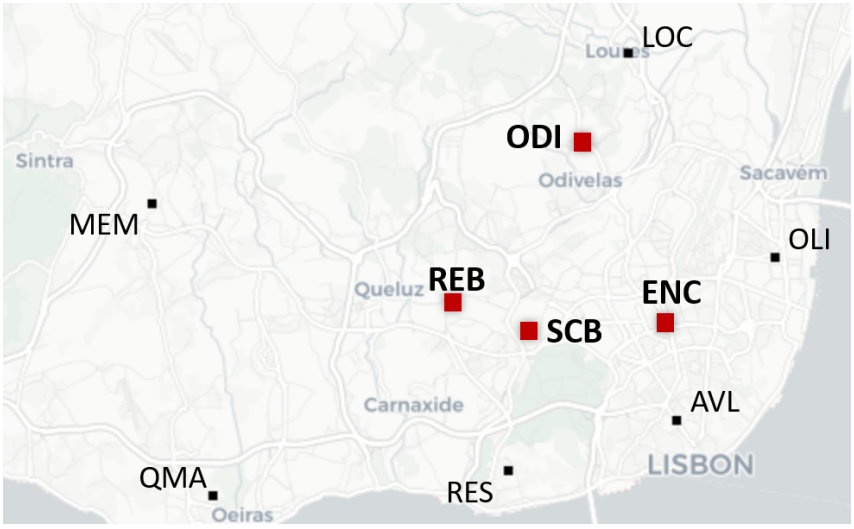

Figure 3.6: Stations in the greater area of Lisbon used in this work.

Variograms can only be modelled if datasets have a relative high number of points and spatial density, which is not the case in this work, in which observation points are only 10 and sparsely distributed [30]. Cross validation was applied to decide which is the best Kriging variogram for this work selected data.

3.2.3 FBN Implementation

FBNs have several implementation details which are user defined.

Antecedent and Consequent Areas

First, the amount of antecedent and consequent areas that constitute the network, and which variables are associated to each area, must be defined. The dataset PM10 values and its coordinates are the variables used. The interpolation objective is to model a value for the PM concentration at a specific location, with a respective pair of coordinates. Therefore, there are two independent variables, represented as antecedent areas - latitude and longitude - and one unknown value, represented as consequent area - PM10 concentration.

Variables Scale Conversion

Since each area only receives inputs between 0 and 1, the values of all used variables (latitude, longitude and PM10 concentration) need to be converted to the same scale, and a maximum value needs to be defined, to represent 100% ratio of its corresponding area.

In the case of the coordinates, delimiting values were represented by the outline of the most exterior stations.

The station at the extreme west location is Mem Martins (MEM) whose longitude is -9.347222, and the extreme east station is Olivais (OLI) with a longitude of -9.108056. Therefore, the longitude defined value for 0 is -9.35 and for 1 is -9.1.

The same was done for the latitude. The southest station is Quinta do Marquês (QMA), which has latitude equal to 38.6975, and the northest is Loures Centro (LOC) whose latitude is 38.829722. Therefore, the defined values for 0 and 1 for the antecedent area corresponding to the latitude coordinate are 38.69 and 38.83 respectively.

With these settings it is possible to define the spatial grid presented in Figure 3.7.

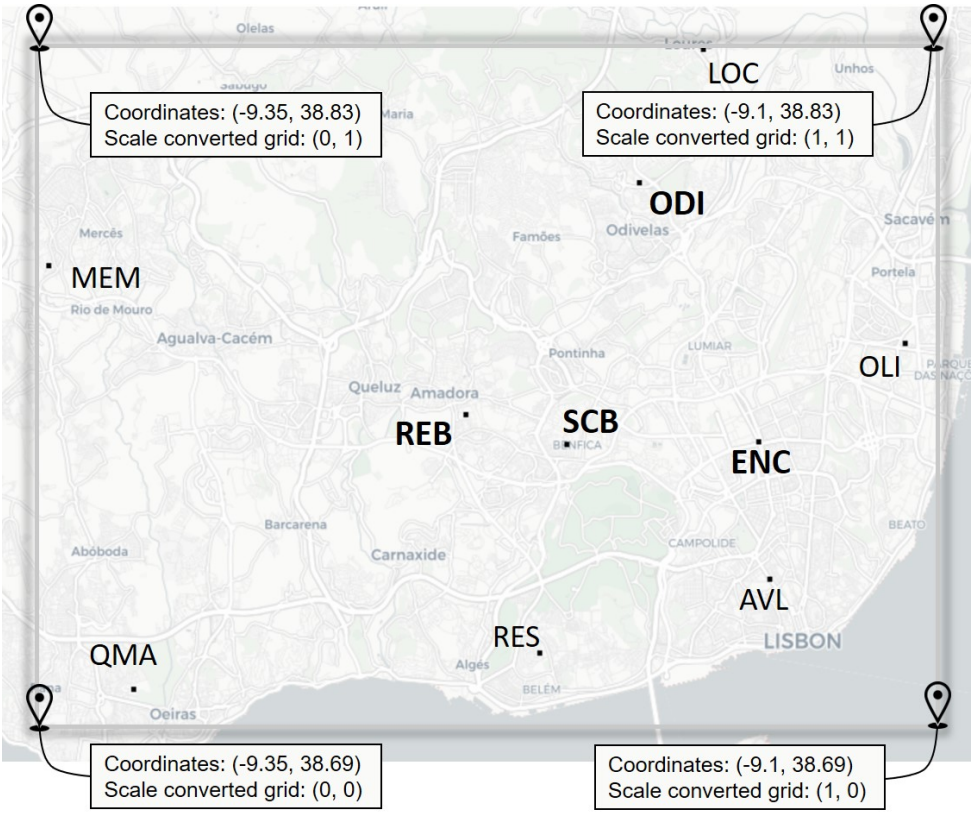

Figure 3.7: Spatial grid used for FBN coordinate scale conversion.

Finally, for the PM10 concentration values, the minimum value was 0 µg/m^3 and the maximum value was 175 µg/m^3, which is above the highest registered value in the filtered 5 year dataset.

The equations from Equation (3.1) to Equation (3.3) represent the scale transformations used for each variables, longitude, latitude and PM10 concentration.

$$Lon_{val} = \frac{Lon_{station} - Lon_{min}}{Lon_{max} - Lon_{min}} \tag{3.1}$$

$$Lat_{val} = \frac{Lat_{station} - Lat_{min}}{Lat_{max} - Lat_{min}} \tag{3.2}$$

$$PM10_{val} = \frac{PM_{observed}}{175} \tag{3.3}$$

Parametrization

In order to better explain and understand FBN performance in this work application, several parametrizations were tested. The user defined parameters which were tested are the number of antecedent and consequent areas in the network, the number of testing and training epochs, the granularity and the number of samples.

Training and testing epochs were tested on how its increase, independently of the network size, influences results. Furthermore, net size along with number of samples and granularity were also tested to conclude how these parameters can provide a better accuracy against time of execution through cross validation in the selected dataset.

In order to analyse how resolution of data influences FBNs interpolation performance results, these were tested with a dataset of 15 minutes intervals values measured at monitoring stations. This dataset was provided by CCDR-LVT, and contained only measures from the stations AVL, ENC, OLI, RES and SCB.

Additionally, FBNs were tested with data from four 15 minute sets of values per hour, to infer an average for that hour at the location of each station. This was different than previous tests, in which only hourly averages were used to train and were infered by FBNs.

Implementation

The implementation of FBN was the result of previous works [27]. This implementation was migrated from Python 2 to Python 3, and code was additionally developed to insert every training and testing rule in the networks directly from an excel dataset. This enables the interpolation of several datasets in only one execution.

3.2.4 Algorithm Performance Evaluation

In order to assess the results of the cross validation approach, the statistical error parameters presented in equations (3.4) and (3.5) were considered.

$$MAE = \frac{\sum_{i=1}^{n} |M_i - O_i|}{n} \tag{3.4}$$

$$RMSE = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(M_i - O_i)^2} \tag{3.5}$$

Where M_i represents the modelled value, O_i represents the observed value and n is the total of inferences.

Mean Absolute Error (MAE) measures the average magnitude of errors in a set of predictions. Relative Mean Square Error (RMSE) is a quadratic parameter which also measures the average magnitude of the error. Both express average model prediction error in units of the variable of interest, which for PM10 concentration are µg/m^3. Regarding RMSE, since errors are squared before they are averaged, increased weight is given to large errors.

3.3 System for Online Data Visualization

The system for online data visualization is composed of a database, an html, css and javascript apache website with a javascript library which provides custom map web visualization and rendering, Mapbox GL JS.

3.3.1 Database

A MongoDB database was developed to store the values obtained by monitoring systems. Depending on the performance of the developed low-cost system, several of its replicas could be placed around Lisbon, in locations where there are no other monitoring stations, in order to complement and increase the density of data for the city of Lisbon.

MongoDB was the database technology chosen for this specific application, due to its high scalability and flexibity. It is a NoSQL type of database, which stores data in flexible, JSON-like documents. It is a distributed database at its core, therefore it has high availability through built-in replication, and horizontal scalability with native sharding, with end-to-end security. Each database contains collections, which in turn contain documents. Documents are filled with fields and the size and content of documents in the same collection can vary.

The database used in this work only has one collection, in which every measure from the system is saved. Despite the non-enforcement of document structure in the database, for the purpose of this work, each valid document must have date, location, sensor ID and pm10 value as valid fields, and the date field along with the sensor ID can not be repeated in different documents.

3.3.2 Mapbox GL JS

Mapbox is a provider of custom online maps for websites and applications. The company was founded in 2010 and has contributed to, and created, several tools which let developers use custom maps in their own platforms. It has emerged as a response to the limited choice offered by map providers such as Google Maps. Some of these tools are open source mapping libraries and applications, such as the Mapbox GL JS JavaScript library, the MBTiles specification, the TileMill cartography IDE, the Leaflet JavaScript library and the CartoCSS map styling language and parser [74].

Mapbox GL JS is a JavaScript library which uses WebGL to render interactive maps from vector tiles and Mapbox styles, and it was used in this work to represent the results of the real-time spatial interpolation PM10 values.

Its online map implementations are widely customizable and allow the visualization of a wide variety of data representations in the chosen map, such as polygons, icons, buildings, polygons, 3D models and animations.

In this work, extrusion of polygons was used to represent the levels of estimated PM10 concentration values, whose height and color is determined by the pollution value in each point of the map.

For the representation of these polygons, a grid was defined, which is delimited by the coordinates of the Lisbon monitoring station networks which are on the edges of the city. This is a convex hull grid, with a resolution of 100 m^2. For these extrusions to be included, a geojson file was created, with the polygons of every element in the defined grid. A python script is running in the background which updates the values at each station every hour, performs a new interpolation with the updated values, and updates the geojson file accordingly.

Regarding the javascript implementation, a Map object, which represents the whole map visualization, is defined in the main website file. In the javascript code of the webpage, the geojson file, which should be saved in a local directory, is accessed, and then its respective extrusions are inserted in the Map object, which is rendered when users access the webpage.

4

Results

Contents

4.1 NB-IoT PM10 Monitoring System 49

4.2 SIMs Testing 53

4.3 Online Data Visualization Web Application 57

4.4 Discussion 58

In this chapter, the results regarding the overall developed system for air quality monitoring and visualization with NB-IoT will be presented, as well as the performance of the considered SIMs. It will be divided into four sections, the PM10 Monitoring System experiment, SIMs performance, Online Data Visualization Web Application and finally Discussion. In this final section, the obtained results are addressed and analyzed in terms of what substantial inferences can be made regarding the experiments of this work.

For a better understanding of the results, it is important to note how PM10 concentrations affect air quality. In Table 4.1 one can see the official classification of the PM10 concentration values [9].

Table 4.1: Classification of PM10 concentration values.

Classification	PM10 concentration interval (µg/m^3)
Very Good	0 - 20
Good	21 - 35
Medium	36 - 50
Weak	51 - 100
Bad	101 - 1200

It can be observed that the smallest classification interval corresponds to a 14 µg/m^3 variation, for both Good and Medium intervals. This value is useful to assess overall error results in relation to official metrics, for both the assessment of SIMs and the errors of measures between the developed system and reference monitors.

4.1 NB-IoT PM10 Monitoring System

This section objective is to describe the developed monitoring system performance and measures in comparison to the reference monitors, to later assess how it may integrate existing reference monitoring networks, which lack geographical density. Furthermore, the state and performance of the NB-IoT technology during the period of its use in this work is characterized.

4.1.1 NB-IoT Coverage in Lisbon

In Table 4.2 are presented the weeks during which the developed system was placed in a CCDR-LVT station. Colored cells represent the location where the developed sensor was placed during each week. In weeks colored in green, consistent measures were taken since there was NB-IoT coverage. On the other hand, in red colored weeks it was impossible to get consistent coverage and consequently no measures were obtained.

Table 4.2: Weeks during which the developed system was placed next to a CCDR-LVT monitoring container.

	July		August				September			
Week	3	4	1	2	3	4	1	2	3	4
AVL										
ENC										

4.1.2 Evaluation of Obtained Measures

From 13 to 26 September, measures obtained for each 15 minute interval were consistent and valid. The resulting data characteristics from these during this period are presented in Table 4.3.

Table 4.3: Data obtained during the measurement period.

Full days	Obtained measures	Resulting hourly averages
12	1195	298

Since data available from the reference monitor only corresponded to hourly averages, these were also calculated for the data obtained from the developed sensor, for comparison purposes. It is important to note that the reference monitor averages available online [9] are made for each hour with all the 15 minute spaced measures from the previous hour. Therefore, the same process was applied to calculate the developed system averages.

Statistics, such as the mean value, standard deviation and maximum and minimum values, regarding both obtained data sets were calculated and are presented in Table 4.4.

Table 4.4: Results of the measurement period, in ($\mu g/m^3$).

	Mean value	Standard Deviation	Maximum value	Minimum value
Reference Monitor	27.88	8.78	55.00	8.00
Developed System	15.04	10.64	56.80	2.49

In Figure 4.1, simultaneous measures of both sensors can be observed, for the whole time of placement of the developed system, along with the average RH values for the city of Lisbon.

Several tendencies can be observed. The PMS5003 sensor often obtained lower measures. However, sometimes it peaked, surpassing the MP101M measures. This only happens for high values of RH. When the RH is decreasing, there is often less error between sensors.

The days 14, 15, 21 and 22 of September corresponded to weekend days, which means there is less car traffic in AVL. No meaningful tendencies were observed in these days, despite the average PM10 concentration being slightly lower in both sensors.

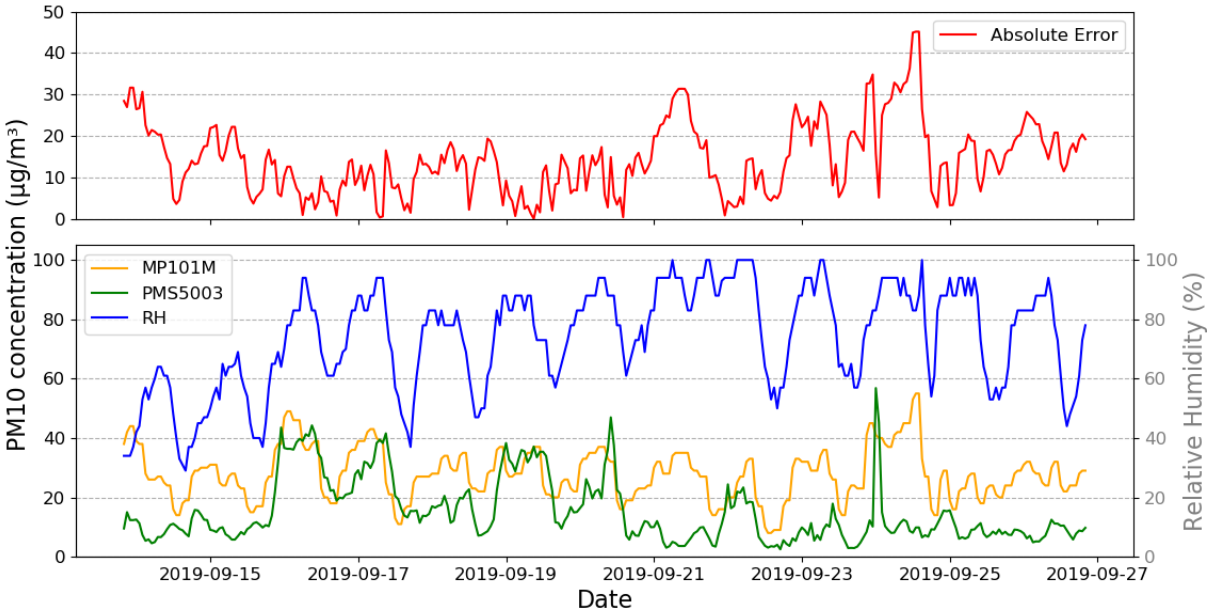

Figure 4.1: Time series of the PM10 concentrations reported by the sensors and corresponding RH.

The average absolute error between the measures of the different sensors is 14.40 µg/m^3. Furthermore, the scatter diagram in Figure 4.2 also highlights that the developed system measured values are predominantly lower than the reference measures.

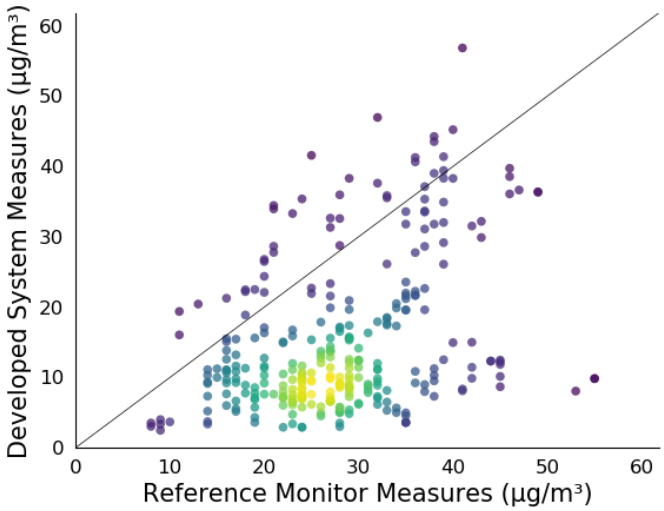

Figure 4.2: Scatter diagram between the developed system and the reference monitor obtained measures.

In the comparison of time series data, it is important to assess not only the error statistics between datasets, but also their linearity. In order to identify correlations between different sensor readings, the Pearson coefficient was used. As explained in Section 3.2.1, the Pearson coefficient represents a measure of the linear correlation between two variables. The resulting coefficient varies from 1 to

-1, where 1 represents total positive linear correlation, 0 represents no linear correlation, and -1 total negative linear correlation.

The Pearson coefficient was calculated for every full day of the time series data, during the developed sensor placement. These and the overall Pearson coefficient are presented in Table 4.5.

Table 4.5: Pearson coefficients calculated for the full days of sensor positioning.

Day of placement (September of 2019)												Overall
14	15	16	17	18	19	20	21	22	23	24	25	
0.22	0.57	0.68	0.60	0.75	0.62	0.59	-0.23	0.57	0.34	0.01	-0.37	0.39

The highest correlation coefficient corresponded to the fifth day of placement, 18 of September, in which the value of the coefficient was 0.75. The day with the lowest correlation values was the twelfth day, 25 of September, with a coefficient of -0.37. In terms of average errors, the day where higher absolute error was observed was 24 of September, with an error of 23.52 µg/m^3, and the day with the lowest was 19 of September, with an error of 6.87 µg/m^3. The measures corresponding to these days, as well as respective RH, are presented individually in Figure 4.3.

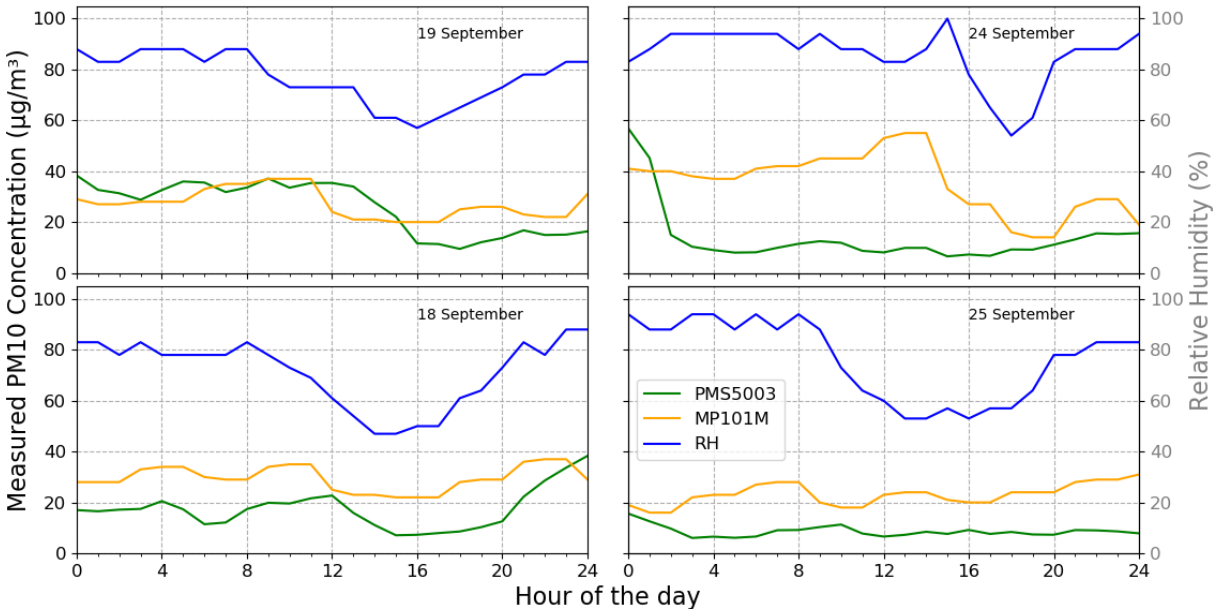

Figure 4.3: Measures of the days with the highest and lowest absolute errors and Pearson coefficients.

4.2 SIMs Testing

In this section, SIM parametrization and performance results are presented. In the parametrization phase, the main objective is to define the best parameters for every interpolation algorithm, with main focus on FBN, which is the less documented algorithm. In both phases, error statistics are calculated in order to define whether there is an algorithm which performance stands out from the remaining.

4.2.1 SIM Parametrization

The used algorithms which have user defined parameters are IDW, OK and FBNs. In this section these parameters are tested in what regards error statistics in the 5 year dataset.

In IDW, the only user defined parameter is the power function. Through the cross validation approach application in the dataset, the power function with the best error statistics was chosen. Values of p from 0 to 5 were tested. Results are presented in Table 4.6, and the value selected, which has the lowest overall MAE and RMSE values is $p = 0$.

Table 4.6: IDW power function comparison results.

p	ENC		ODI		REB		SCB	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
0	4.57	6.34	4.97	6.59	5.79	7.37	11.15	14.88
1	4.74	6.25	4.97	6.62	7.19	9.04	10.73	14.34
2	5.69	7.38	5.15	6.88	9.40	11.94	11.06	14.64
3	6.80	8.79	5.61	7.52	11.11	14.33	11.80	15.42
4	7.76	10.06	6.17	8.26	11.92	15.49	12.47	16.15
5	8.54	11.09	6.60	8.86	12.24	15.94	12.90	16.62

The same approach was applied in the parameter assessment of the remaining algorithms.

In OK, the variogram is the main user defined parameter. As stated in Section 3.2.2, the dataset used in this work's problem has a very low number of data points, which are sparsely located in space. This problem makes it not viable to define a custom variogram for the OK interpolation. Therefore, several pre-defined variograms were tested, in order to identify which was the most suitable for this specific interpolation problem.

The obtained results are presented in Table 4.7. The variogram model which has the lowest overall MAE and RMSE values is the linear model.

In what regards FBN interpolation, an higher number of parameters need to be addressed, namely both testing and training epochs, the network size, the number of samples made by each neuron and

Table 4.7: OK variogram comparison results.

Variogram	ENC		ODI		REB		SCB	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Linear	4.95	6.68	5.03	6.69	5.88	7.52	10.89	14.51
Gaussian	5.96	8.02	5.38	7.29	8.93	12.23	11.21	14.69
Exponential	5.42	7.24	5.14	6.87	7.28	9.38	10.94	14.49
Spherical	5.72	7.66	5.25	7.04	8.24	10.77	11.19	14.69
Power	4.97	6.68	5.08	6.77	6.12	7.89	10.88	14.49
Hole-effect	4.77	6.62	5.03	6.65	5.95	7.57	11.36	15.08

the granularity of the memories of each consequent neuron.

Tests were made with several samples, with 2000 iterations, from the dataset, with the objective to assess the set of training and testing epochs, and the set of the net size, number of samples and granularity, independently, as can be observed in Table 4.8.

Table 4.8: FBN Parametrization assessment results.

Training Epochs	Testing Epochs	Net Size	Samples	Granularity	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)
10	10	15	4	2	13.99	18.35
50	50	15	4	2	13.63	17.96
10	10	75	9	3	11.22	14.80
50	50	75	9	3	10.98	14.54
10	10	125	16	4	12.94	16.91
50	50	125	16	4	12.05	15.92

From further studying and experimenting with FBNs, it was observed that:

- Performance increased with the number of testing and training epochs;
- Similarly to the observed in other studies [27], the number of samples should be substantially lower than the net size;
- When very low granularity values are used, consequent neurons will have very small memory tables. Therefore, the information sampled in sampling phases, in both learning and testing epochs, will have insufficient resolution;

- For high granularity values, the number of testing and learning epochs needs to be higher, than with values close to the square root of number of samples.
- As also pointed in previous studies, FBNs advantages lie in their ability to obtain very good results without extensive parameterization efforts [27].

The chosen parameters for the FBN interpolation, accounting the previous results, are presented in Table 4.9.

Table 4.9: Choice of parameters for FBNs used in this work.

Training Epochs	Testing Epochs	Net Size	Samples	Granularity
50	50	75	9	3

FBN interpolation performance was also tested using the dataset with measures corresponding to 15 minute intervals, as referred in Section 3.2.3. This test main objective is to analyze how resolution of data influences FBNs interpolation performance.

The resulting error statistics are presented in Table 4.10.

Table 4.10: FBN 15 minute improvements.

	AVL		ENC		OLI		RES		SCB	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Regular FBN	8.36	10.99	7.18	9.52	9.72	11.97	6.05	9.38	6.53	9.68
15min FBN	9.42	12.49	8.87	11.87	10.77	13.35	6.72	10.41	6.50	10.17

In the regular FBN execution these values were 7.57 and 10.57, which are better than the 15 minute implementation, which were 8.46 and 11.71.

4.2.2 SIM Comparison

Every algorithm considered so far in this work was tested with the cross validation approach in the 5 year dataset.

Results regarding the error statistical parameters MAE and RMSE of every algorithm are presented in table Table 4.11, for a total of 43 956 single inferences each.

The algorithms with the overall lower MAE and RMSE results were IDW, with 6.42 µg/m^3 and 9.25 µg/m^3, respectively, and OK, with 6.42 µg/m^3 and 9.23 µg/m^3. NN with 11.03 µg/m^3 and 14.66 µg/m^3, had the worse results, followed by LI, with 8.28 µg/m^3 and 11.36 µg/m^3. FBN results were 7.55 µg/m^3

Table 4.11: Model comparison results.

	ENC		ODI		REB		SCB	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
IDW	4.65	6.52	4.89	6.48	5.46	6.97	11.52	15.31
OK	4.67	6.52	4.89	6.49	5.55	7.06	11.41	15.18
FBN	5.82	7.62	6.36	8.41	7.55	9.62	11.08	14.75
Linear	5.90	7.74	6.59	9.46	10.11	13.11	11.04	14.51
NearestN	11.20	14.63	7.35	9.82	12.58	16.36	13.45	17.22

and 10.24 µg/m^3. Density scatter diagrams and histograms of the IDW, OK and FBN algorithms can be observed in Figure 4.4 and Figure 4.5.

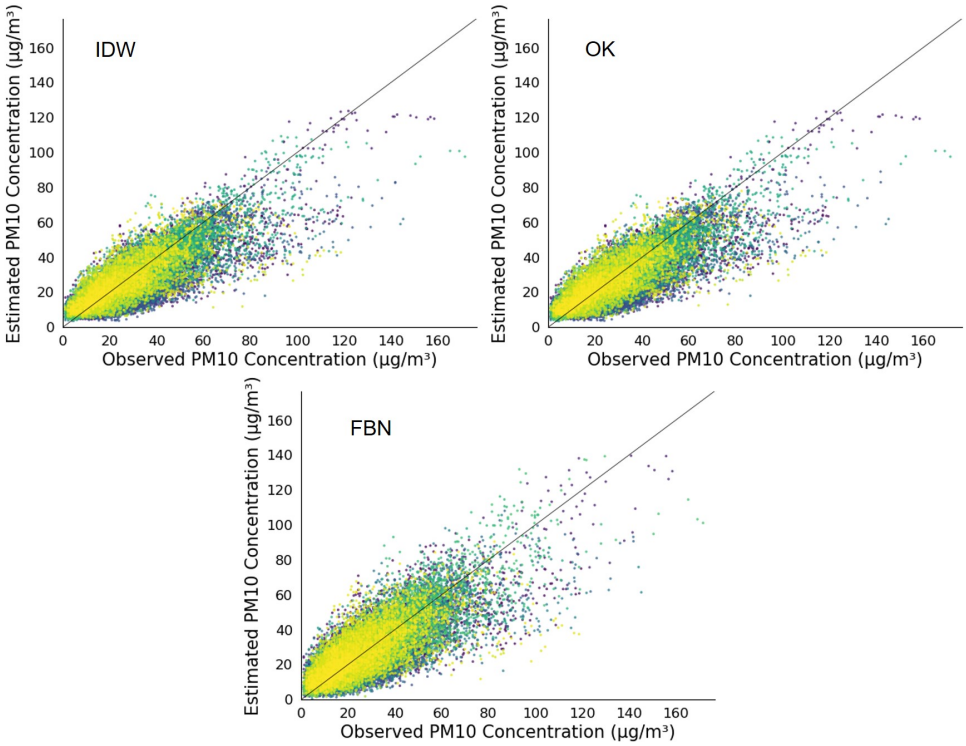

Figure 4.4: Scatter diagrams obtained from 43 956 inferences for IDW, OK and FBN, respectively.

It should additionally be noted that FBNs execution times of the obtained results took approximately 9 hours, while every remaining algorithm only took from 5 to 10 minutes interpolating through the entire dataset.

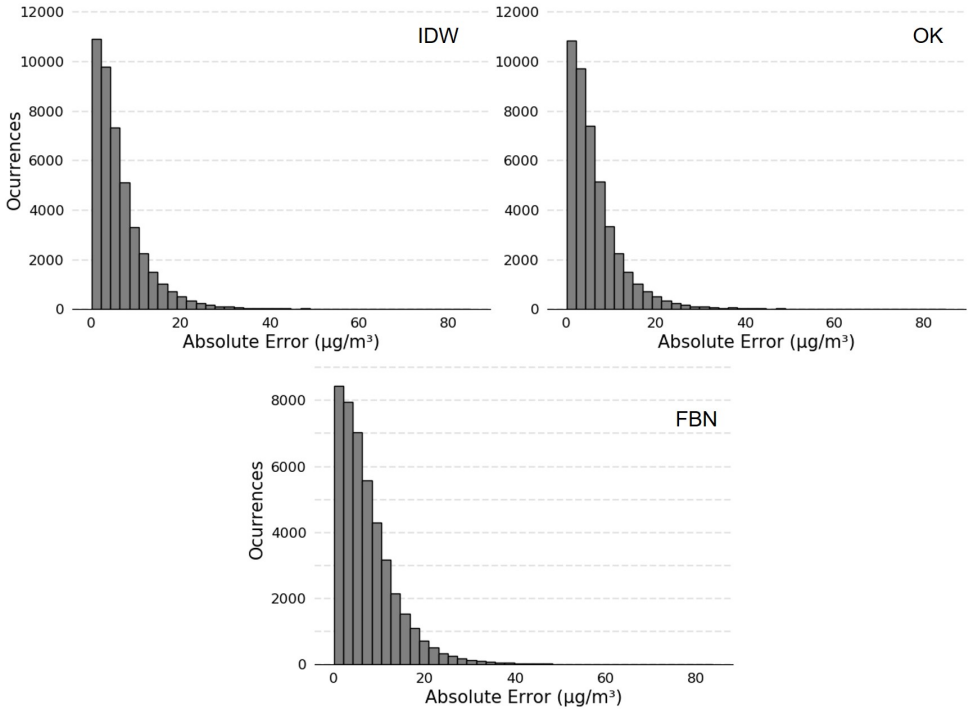

Figure 4.5: Histograms obtained from 43 956 inferences for IDW, OK and FBN, respectively.

4.3 Online Data Visualization Web Application

The developed PM10 visualization application was created to present a tool to visualize tested geographical interpolation algorithms and also to constitute a prototype for a platform of live online visualization of interpolated air quality related phenomena.

The resulting prototype is an application with PM10 interpolated data for the city of Lisbon. It has a resolution of 100 meters, both in color and in height of the pollution surface. It provides the possibility to hover over to know estimated PM10 concentration values in any location inside the outline defined by the CCDR-LVT monitoring network stations.

The application design is focused in the user experience, and allows the user to zoom in and out, and change both pitch and bearing of the map.

The final result of the developed web application is presented in Figure 4.6 and is available at `http://138.68.165.208/`.

In Table 4.12 are presented the access times of some of the platforms for air pollution visualization referenced in this work, on comparison to the developed one. These results were provided from tests from Pingdom tool for web page performance assessment [75].

Figure 4.6: Developed web application final result, with live data from QualAr, interpolated with OK.

Table 4.12: Access times of various air quality visualization web pages.

Web Page	Access Time (ms)	Requests	Page Size (MB)
138.68.165.208 (developed application)	207	44	7.5
airindex.eea.europa.eu	2580	77	3.8
qualar.apambiente.pt	3400	180	9.1
breezometer.com/air-quality-map/city/lisbon	3629	184	2.2

4.4 Discussion

Results obtained and presented in this chapter are discussed in this section.

4.4.1 Developed PM10 Monitoring System

Regarding the NB-IoT coverage in Portugal, from both MEO and NOS, namely in the capital Lisbon, it could be concluded that it is not good enough for professional applications at the time of this work. During the 10 weeks of placement of the developed system, in different locations in the city of Lisbon, there were only 2 weeks when it was possible to get consistent coverage.

According to its specifications, overviewed in Section 2.4.4, NB-IoT has one of the widest coverages between IoT technologies. This was not observed in this work. It might have happened because NB-IoT is still in early implementation stages in Portugal.

These results demonstrate that the developed sensor, at the time of this work, from the network

coverage perspective, can not be employed neither integrate current official monitoring networks.

Before assessing the used low-cost sensor performance, it is important to note that the obtained data only corresponds to a period of approximately two weeks as presented in Table 4.2, which represents a small sample size in comparison to what was expected, due to the NB-IoT network coverage constraints.

During the sensor placement and comparison period, the PMS5003 sensor often obtained lower measures. However, sometimes it peaked, surpassing the MP101M measures. This only happened for high values of RH. It additionally can be observed in Figure 4.1 that when the RH is decreasing is when often less error is observed between the sensors.

The average absolute error between the measures of the different sensors (14.40 µg/m^3) is slightly above the smallest scale division in Table 4.1. In the scatter diagram of Figure 4.2 it can be observed that the developed system measured values are predominantly lower than the reference measures. The daily pearson correlation coefficients (Table 4.5) showed inconsistency in the linearity between sensor readings.

PMS5003 sensor manufacturers claim that the sensor presents high consistency in the temperatures between 10 and 40ºC [20]. It was summer season during this experiment, and the temperature values never exceeded positively nor negatively the above mentioned interval. However, the reference monitor has a controlled sampling flow rate of 1 m^3/h, as well as a temperature-regulated sampling tube. This does not happen in the PMS5003, in which neither air flow nor its temperature are controlled. Furthermore, it is important to note that the low-cost sensor was placed in the outside of the monitoring container, while the reference monitor was placed on the inside.

The only rain occurrences during the placement periods happened in the day 21 of September. It is the day with the second highest absolute error between readings, and PMS5003 measures were very low in comparison to the MP101M.

In a study by Zheng et al., in 2018, a similar sensor model (PMS3003) was extensively tested against professional reference monitors and it was observed that while temperature took a negligible part in the difference observed between sensors, RH had a considerable influence in the deviation observed in the low-cost sensor [12]. This can be the one cause for the high observed error and linearity inconsistencies in this work experiment.

Humidity can have a negative effect on electronic circuits, such as the one in the developed system. However, no recurrent effects were observed due to humidity peaks specifically in the PMS5003 readings.

The registered high humidity intervals correspond to night and dawn periods. Similar to what happened in a study by Jayaratne et al., in 2018, in these periods, measures are often higher than in the rest of the time. The increase of measured values when there is high humidity is expected in the PMS5003 due to its documented humidity sensitivity, however it suggests that the MP101M is also influenced by

high RH.

In the same study, it was stated that above 75% humidity, marked effects are seen in low-cost sensors due to air particle size growth by deliquescence, specially if these do not have drying facilities at the sample inlets [76]. However, in this work, no effect in the PMS5003 was observed in high humidity periods. There is a substantial increase in measured values, but coincident with the increase reported by the MP101M, and there are undervalued measurements during the rain periods.

The days 14, 15, 21 and 22 of September corresponded to weekend days, which often means there is less car traffic in AVL. No meaningful tendencies were observed in these days, despite the average PM10 concentration being slightly lower in both sensors.

During the day, MP101M readings are mostly higher than PMS5003's. According to Jayaratne et al., in the same study [76], this may happen because the size of traffic particles are below the minimum detectable size limit of low-cost sensors, which for the PMS5003 is 0.3 μm (300 nm). As presented in a study by Li et al., in 2018, where the size-resolved mass concentration of volatile and non-volatile components of vehicle emitted particles is studied, the diameter of particles emitted by vehicles was mostly below 300 nm, which corroborates the fact these are not accordingly measured by the PMS5003 [77].

The objective of this experiment was to find out whether it is possible to define a correction factor for the low-cost sensors, through which they could provide PM10 readings with very low error in comparison with reference monitors. Results show that the absolute error between sensors is not linear neither correlated with the absolute value of RH. Therefore appropriate correction factors are not possible to derive with the obtained data.

Due to the numerous disadvantages observed in this work, it can be concluded, with this work data, that low-cost sensors are not suited for regulatory applications, such as assess if the air quality meets stipulated guidelines for particle pollution. However, with previous study of the placement location, such as average particle size and overall RH, and the deduction of a fitting correction factor to that specific location, it could be used for informative purposes, for example in interpolation platforms, to provide a better resolution to monitoring networks which currently suffer from sparsity.

In mentioned studies concerning PMS low-cost sensor calibration, it was concluded that appropriate calibration models using dynamic adjustments for meteorological parameters are an essential prerequisite for these sensors to achieve high accuracy and precision in the ambient monitoring and that after proper calibration, low-cost PM sensors can be promising monitors for PM sensor network development and to complement the existing networks, without specifying its integration in official networks [12].

4.4.2 SIMs Assessment

The algorithms with the best overall performance were IDW and OK, followed by FBNs, LI and lastly NN.

In Table 3.4, correlation coefficients between every station of the dataset was calculated. It can be observed that the stations with the highest correlation coefficients between them were relatively close in terms of geographical location, despite an overall big correlation observed between almost every station.

The high performance of IDW with $p = 0$, which means that distance between data points is not considered for the calculation of the interpolated values, suggests that the distance correlation between stations is low.

Air dispersion and movement in urban environments is an environmental problem of high complexity, due to its influence by many factors. In the last 5 years, in the city of Lisbon, the dataset reveals the city to have an overall low concentration of PM10, with maximum mean value for the whole dataset, between every station, being 34.22 ug/m3 in the station AVL, which is a value classified as *Good*.

The fact that these SIMs were tested for periods where no big pollution phenomena affected the city, could degrade the sense of distance correlation between the stations in the city, because the collected data was predominantly constant and similar between stations.

Other environmental factors can influence the correlation of air quality throughout the city, such as the wind direction, the temperature, humidity, the altitude of the interpolation point and the traffic density and pollution, as well as the presence of buildings and other geological and geographical barriers to airflow.

In what regards FBNs performance, no conclusions can be obtained towards its performance regarding multi dimensional spatial interpolation problems due to the low distance correlation between the data points of the performed experiments. However it can be concluded that FBNs are not suitable for problems which require real time interpolations with fast computations and construction of interpolation surfaces, due to the long computing times observed in this work.

4.4.3 Web Visualization Application

The developed application resulted in a user-friendly platform, with the presentation of live PM10 concentration data with high resolution in the area covered by the CCDR-LVT monitoring stations network in Lisbon. Since results obtained from the low-cost sensors reveal these sensors lack performance, without specific calibration corrections, these were not included in the network used for the spatial interpolation in the application.

The application has good performance and is flexible from a development point of view. It can easily be extended to support several other air polluting gases concentration or indicators. The same structure could additionally be applied in other geographical interpolation fields.

In what concerns web page performance, the developed application has an extremely lower access time than other mentioned web pages, as well as a less number of requests. This is expected since it does not use any kind of back end processes, besides the ones of Mapbox GL JS, and is only composed

of a map with interpolated data, while other web pages have several more web elements and components which can cause delays in access times. High page size was also observed and expected, since no component compression was built into the application.

In comparison with current governmental web applications available, the developed application has higher resolution and provides better user experience, despite still not covering large areas such as whole countries. Furthermore, it is already prepared for a future state of official monitoring networks with increased density of monitoring stations and an interpolation representation which accuracy is expected to rise with the increase of data points.

As a product, there are already companies with applications which outperform it, as seen in Section 2.6 of this work, being the most evident Breezometer.

Breezometer presents data regarding numerous air polluting gases and indicators with a resolution of 500 meters [66], and worldwide coverage. However, due to its proprietary characteristics, its used air modelling algorithms and used tools are not documented and consequently not available for assessment.

It is important to note that the characteristics enumerated do not constitute disadvantages of the developed application, but rather a comparison and position of the developed prototype with the state of the art of the current web visualization applications for air quality pollution.

It can be concluded that the developed platform could bring awareness to the citizens in what regards their cities air quality, in high resolution, and constitute a replacement for current government platforms.

5

Conclusion

Contents

5.1 Conclusion 65

5.2 Future Work 66

The main objective of the work developed in this master thesis is the analysis of the current state of the information available publicly regarding air quality, namely PM10 concentration data, in the city of Lisbon, and to experiment with different technologies which could improve it. This was done through the development of a prototype for a low-cost PM10 sensor node with the use of the NB-IoT technology, the creation of a live PM10 concentration visualization application, and the assessment of various spatial interpolation algorithms in the inference of PM10 concentration in the city of Lisbon, with the limited data points available of the CCDR-LVT air quality monitoring network. This work conclusions solidify knowledge on the application of every used technology, in the field of air quality monitoring, interpolation and visualization.

5.1 Conclusion

The developed low-cost sensor was successfully deployed in AVL station for two weeks. Measures were taken and compared with the simultaneous reference AVL PM10 monitor. Results showed an error of 14.40 µg/m^3 between the two, which is slightly above the smallest scale division of PM10 classification. Inconsistent linearity was observed during the whole period, through the calculated daily Pearson correlation coefficients. Finally, it was concluded that the sensor is not able to integrate current official monitoring networks, and further experiments should be made to better understand these low-cost sensors calibration adjustments.

The spatial interpolation algorithms tested, with data constituting 5 years of Lisbon's monitoring network measurements, from 2013 to 2017, revealed that there is not an high distance correlation between the stations in the network, in terms of geographical location influence. These were used only with geographic coordinates, and pollution in Lisbon was considered low in the considered 5 years, with very sparse data points, specially accounting the present urban topology, which might constitute reasons for why IDW with $p = 0$, was one of the best performing algorithms. Finally, the developed online visualization platform resulted in a web application which can provide PM10 data with finer resolution than other current available platforms, even though it only covers the city of Lisbon, and in an approach which can be used by further studies and platforms to represent geographical interpolation phenomenons online.

As an overall conclusion to this work, the current state of the data available publicly regarding air quality was studied and assessed and several experiments were made in other to evaluate the application of new technologies, and the creation of prototypes, which could improve it. Relevant results were obtained regarding the usage and performance of every technology used and applied in this work. Finally, information was gathered for further study of the phenomenon of spatial interpolation of air quality, its measurement and its visualization.

5.2 Future Work

In a study using NB-IoT, several tests and an evaluation of the study area coverage should be made previously to its application. These tests should be made in the field and with several different mobile operators and different NB-IoT chipsets, in order to deduce coverage constraints and hardware limitations.

Low-cost PM monitoring sensors, should be studied beforehand in controlled environments. Various tests should be made with different conditions in order to evaluate the deviations between measures as an answer to real world conditions.

Ideally, with a very well understanding of this sensor properties, and a calibration function which could minimize the error to extremely low values, for each specific placement location, this type of sensors could integrate air quality monitoring networks, but only to further increase their spatial density and the resolution at which air quality could be visualized. This would ultimately improve researchers understanding of air quality phenomenons, and propel research progress in related fields of study.

Spatial algorithms tested in this work could be studied and complemented with air dispersion models. Data could be gathered for certain long periods of time for variables such as traffic, wind direction, temperature, altitude and relative humidity, and could be incorporated into machine learning algorithms.

Regarding FBNs, its implementation could be parallelized, which could drastically optimize its performance, due to the parallel nature of its neuron sampling operations. It could be further studied regarding spatial geographical interpolation, with the integration of additional variables to the problem.

The web visualization tool developed in this work, as already pointed out in Section 4.3 of this work, could be used in the future as a visualization tool in different geographical fields. It could also be extended in order to include other air quality pollution gases and indicators.

The developed prototype could also be extended into a web application, since it is constituted predominantly by web components, through the development of an adequate application programming interface, and the deployment with a mobile app development technology, such as Ionic, which allows the development of mobile applications using web technologies such as CSS and HTML.

Bibliography

[1] WHO, "9 out of 10 people worldwide breathe polluted air, but more countries are taking action," https://www.who.int/news-room/detail/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action, last accessed on August, 2019.

[2] ——, "More than 90% of the world's children breathe toxic air every day," https://www.who.int/news-room/detail/29-10-2018-more-than-90-of-the-world's-children-breathe-toxic-air-every-day, last accessed on August, 2019.

[3] GSMA, "Air Quality Monitoring Using IoT and Big Data, A value generation guide for mobile operators," 2018, https://www.gsma.com/iot/wp-content/uploads/2018/02/iot_clean_air_02_18.pdf, last accessed on August, 2019.

[4] A. Russo and A. O. Soares, "Hybrid Model for Urban Air Pollution Forecasting: A Stochastic Spatio-Temporal Approach," *Mathematical Geosciences*, vol. 46, no. 1, pp. 75–93, 2014.

[5] APA, *Evolução da qualidade do ar em Portugal entre 2001 e 2005*, 2008.

[6] D. F. Agência Portuguesa do Ambiente, "Qualar - Base de Dados Online sobre a Qualidade do Ar - Estações Activas," www.qualar.org/index.php?page=4{&}subpage=5{&}x=111{&}y=289, last accessed on August, 2019.

[7] EEA, "Air quality in Europe — 2018 report," pp. 1–64, 2018.

[8] S. S. Amaral, J. A. de Carvalho, M. A. M. Costa, and C. Pinheiro, "An overview of particulate matter measurement instruments," *Atmosphere*, vol. 6, no. 9, pp. 1327–1345, 2015.

[9] QualAr, "Mapa dos índices diários provisórios (não validados)," https://www.qualar1.apambiente.pt/qualar/, last accessed on August, 2019.

[10] Laboratório de Referência do Ambiente, *Manual de Métodos e de Procedimentos Operativos das Redes de Monitorização da Qualidade do Ar - Amostragem e Análise*, 2010.

[11] E. S.A., "Mp101m – continuous, automatic pm10, pm2.5, pm1, tsp particulate monitor," http://www.environnement-sa.com/products-page/en/air-quality-monitoring-en/mp101m-continuous-automatic-pm10-pm2-5-pm1-tsp-particulate-monitor/, last accessed on October, 2019.

[12] T. Zheng, M. H. Bergin, K. K. Johnson, S. N. Tripathi, S. Shirodkar, M. S. Landis, R. Sutaria, and D. E. Carlson, "Field evaluation of low-cost particulate matter sensors in high-and low-concentration environments," *Atmospheric Measurement Techniques*, vol. 11, no. 8, pp. 4823–4846, 2018.

[13] E. A. Wendt, C. W. Quinn, D. D. Miller-Lionberg, J. Tryner, C. L'Orange, B. Ford, A. P. Yalin, J. R. Pierce, S. Jathar, and J. Volckens, "A low-cost monitor for measurement of fine particulate matter and aerosol optical depth. Part 1: Specifications and testing," *Atmospheric Measurement Techniques Discussions*, vol. 30, no. April, pp. 1–18, 2019.

[14] D. Quelhas, "Mobile Sensor Network for Air Quality Mapping," Master's thesis, Instituto Superior Técnico, 2011.

[15] V. Carvalho, "Estação Móvel para Medida da Qualidade do Ar," Master's thesis, Instituto Superior Técnico, 2008.

[16] T. Farias, L. Sousa, J. Dias, and R. Ribeiro, "Proposta de criação do Laboratório de Veículos e Sistemas de Propulsão do Departamento de Engenharia Mecânica," Tech. Rep., 2013.

[17] A. Manikonda, N. Zíková, P. K. Hopke, and A. R. Ferro, "Laboratory assessment of low-cost PM monitors," *Journal of Aerosol Science*, vol. 102, pp. 29–40, 2016.

[18] W. Underground, "Weather underground," https://www.wunderground.com, last accessed on August, 2019.

[19] AQICN, "Gaia Air Quality Monitors - Product Series," https://aqicn.org/products/gaia/, last accessed on August, 2019.

[20] T. Sayahi, A. Butterfield, and K. Kelly, "Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors," *Environ Pollut*, 2018.

[21] J. Kuula, H. Kuuluvainen, T. Rönkkö, J. V. Niemi, E. Saukko, H. Portin, M. Aurela, S. Saarikoski, A. Rostedt, R. Hillamo, and H. Timonen, "Applicability of Optical and Diffusion Charging-Based Particulate Matter Sensors to Urban Air Quality Measurements," *Aerosol and Air Quality Research*, vol. 19, no. 5, pp. 1024–1039, 2019.

[22] A. Q. S. P. E. Center, "Air Quality Egg 2018 Model," http://www.aqmd.gov/aq-spec/product/air-quality-egg-2018-model, last accessed on August, 2019.

[23] A. Stein, “Spatial Interpolation,” *Biometrics*, vol. 50, no. 2, p. 592, 1994.

[24] J. Li, “A review of spatial interpolation methods for environmental scientists,” *Canberra: Geoscience Australia*, vol. 137, 2008.

[25] J. Li and A. D. Heap, “Spatial interpolation methods applied in the environmental sciences: A review,” *Environmental Modelling and Software*, vol. 53, pp. 173–189, 2014.

[26] J. Tomé, “Neural Activation Ratio Based Fuzzy Reasoning,” *1998 IEEE International Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence*, pp. 1217–1222, 1998.

[27] J. A. Tomé and J. P. Carvalho, “Fuzzy Boolean Nets - A nature inspired model for learning and reasoning,” *Fuzzy Sets and Systems*, vol. 253, pp. 1–27, 2014.

[28] J. P. Carvalho, “Rule Capacity in Fuzzy Boolean Networks,” *2002 Annual Meeting of the North American Fuzzy Information Processing Society Proceedings*, pp. 124–128, 2002.

[29] J. P. Carvalho and J. Tomé, “Qualitative optimization of Fuzzy Causal Rule Bases using Fuzzy Boolean Nets,” *Fuzzy Sets and Systems*, vol. 158, no. 17, pp. 1931–1946, 2007.

[30] S. M. P. Mesquita, “Modelação da distribuição espacial da qualidade do ar em lisboa usando sistemas de informação geográfica,” Master’s thesis, Instituto Superior de Estatística e Gestão de Informação da Universidade Nova de Lisboa, 2009.

[31] D. Deligiorgi and K. Philippopoulos, “Spatial Interpolation Methodologies in Urban Air Pollution Modeling: Application for the Greater Area of Metropolitan Athens, Greece,” *Advanced Air Pollution*, 2011.

[32] M. Mohan, S. Bhati, A. Sreenivas, and P. Marrapu, “Performance evaluation of AERMOD and ADMS-Urban for total suspended particulate matter concentrations in Megacity Delhi,” *Aerosol and Air Quality Research*, vol. 11, no. 7, pp. 883–894, 2011.

[33] R. F. Lee and M. Drive, “Comparison of Features and Data Requirements among the CALPUFF, AERMOD, and ADMS Models,” 2013.

[34] Y. Zhang, M. Bocquet, V. Mallet, C. Seigneur, and A. Baklanov, “Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects,” *Atmospheric Environment*, vol. 60, pp. 656–676, 2012.

[35] ——, “Real-time air quality forecasting, part I: History, techniques, and current status,” *Atmospheric Environment*, vol. 60, pp. 632–655, 2012.

[36] F. Funenga, “Public Information System about the Air Quality in Lisbon Electrical and Computer Engineering Examination Commitee,” Master’s thesis, Instituto Superior Técnico, 2013.

[37] M. Tang, X. Wu, P. Agrawal, S. Pongpaichet, and R. Jain, “Integration of Diverse Data Sources for Spatial PM2.5 Data Interpolation,” *IEEE Transactions on Multimedia*, vol. 19, no. 2, pp. 408–417, 2017.

[38] J. Oteros, K.-C. Bergmann, A. Menzel, A. Damialis, C. Traidl-Hoffmann, C. Schmidt-Weber, and J. Buters, “Spatial interpolation of current airborne pollen concentrations where no monitoring exists,” *Atmospheric Environment*, 2018.

[39] L. Makowski, “Mobile Sensor Network for Suspended Particulate Matter Measurements using Kriging Interpolation,” *19th International Conference Computational Problems of Electrical Engineering*, pp. 1–4, 2018.

[40] D. Mishra, P. Goyal, and A. Upadhyay, “Artificial intelligence based approach to forecast PM2.5 during haze episodes: A case study of Delhi, India,” *Atmospheric Environment*, vol. 102, pp. 239–248, 2015.

[41] A. Alimissis, K. Philippopoulos, C. G. Tzanis, and D. Deligiorgi, “Spatial estimation of urban air pollution with the use of artificial neural network models,” *Atmospheric Environment*, vol. 191, pp. 205–213, 2018.

[42] F. Franceschi, M. Cobo, and M. Figueredo, “Discovering relationships and forecasting PM10 and PM2.5 concentrations in Bogotá Colombia, using Artificial Neural Networks, Principal Component Analysis, and k-means clustering,” *Atmospheric Pollution Research*, vol. 9, no. 5, pp. 912–922, 2018.

[43] J. Holler, V. Tsiatsis, C. Mulligan, S. Avesand, S. Karnouskos, and D. Boyle, *From machine-to-machine to the internet of things introduction to a new age of intelligence*, 2014.

[44] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, “A comparative study of LPWAN technologies for large-scale IoT deployment,” *ICT Express*, 2018.

[45] M. Centenaro, A. Zanella, and M. Zorzi, “Long-Range Communications in Unlicensed Bands : The Rising Stars in the IoT and Smart City scenarios,” *IEEE Wireless Communications*, vol. 23, pp. 60–67, 2016.

[46] Y. W. Rashmi Sinha, “A survey on lpwa technology: Lora and nb-iot,” *ICT Express*, vol. 3, pp. 14–21, 2017.

[47] Altice, “NarrowBand-IoT is the future and Altice Portugal are bringing it,” https://www.telecom.pt/en-us/media/noticias/Pages/2018/novembro/NarrowBand-IoT-.aspx, last accessed on August, 2019.

[48] SigFox, “SigFox Coverage,” https://www.sigfox.com/en/coverage, last accessed on August, 2019.

[49] T. T. N. Portugal, “The things network,” https://www.thethingsnetwork.org/country/portugal/, last accessed on August, 2019.

[50] W. Yang, M. Wang, J. Zhang, J. Zou, M. Hua, T. Xia, and X. You, “Narrowband Wireless Access for Low-Power Massive Internet of Things : A Bandwidth Perspective,” *IEEE Wireless Communications*, pp. 2–9, 2017.

[51] LoRaAlliance and LoRa Alliance, *LoRaWAN 1.1 Specification*, 2017, no. 1.1.

[52] R. Ratasuk, N. Mangalvedhe, Y. Zhang, M. Robert, and J. P. Koskinen, “Overview of narrowband IoT in LTE Rel-13,” *2016 IEEE Conference on Standards for Communications and Networking, CSCN 2016*, 2016.

[53] Y. Song, J. Lin, M. Tang, and S. Dong, “An Internet of Energy Things Based on Wireless LPWAN,” *Engineering*, vol. 3, no. 4, pp. 460–466, 2017.

[54] H. Mroue, A. Nasser, S. Hamrioui, B. Parrein, E. Motta-Cruz, and G. Rouyer, “MAC layer-based evaluation of IoT technologies: LoRa, SigFox and NB-IoT,” *2018 IEEE Middle East and North Africa Communications Conference, MENACOMM 2018*, pp. 1–5, 2018.

[55] A. Ikpehai, B. Adebisi, K. M. Rabie, K. Anoh, R. E. Ande, G. S. Member, M. Hammoudeh, H. Gacanin, S. Member, and U. M. Mbanaso, “Low-Power Wide Area Network Technologies For Internet of Things : A Comparative Review,” *IEEE Internet of Things Journal*, 2018.

[56] J. Finnegan and S. Brown, “An Analysis of the Energy Consumption of LPWA-based IoT Devices,” *2018 International Symposium on Networks, Computers and Communications (ISNCC)*, pp. 1–6, 2018.

[57] M. Lauridsen, H. Nguyen, B. Vejlgaard, I. Z. Kovacs, P. Mogensen, and M. Sorensen, “Coverage Comparison of GPRS, NB-IoT, LoRa, and SigFox in a 7800 km Area,” *IEEE Vehicular Technology Conference*, vol. 2017-June, pp. 2–6, 2017.

[58] M. Rahman, A. Rahman, H.-J. Hong, L.-W. Pan, M. Y. Sarwar Uddin, N. Venkatasubramanian, and C.-H. Hsu, “An adaptive IoT platform on budgeted 3G data plans,” *Journal of Systems Architecture*, 2018.

[59] R. Shete and S. Agrawal, “IoT based urban climate monitoring using Raspberry Pi,” *International Conference on Communication and Signal Processing, ICCSP 2016*, pp. 2008–2012, 2016.

[60] C. Balasubramaniyan and D. Manivannan, “IoT enabled Air Quality Monitoring System (AQMS) using Raspberry Pi,” *Indian Journal of Science and Technology*, vol. 9, no. 39, 2016.

[61] S. R. Patil, “Smart Environmental Monitoring through Internet of Things (IoT) using RaspberryPi 3,” *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, pp. 595–600, 2017.

[62] V. Ylläsjärvi, “Overview of Nb-IoT chipset vendor’s offerings,” https://haltian.com/news/nb-iot-chipset-vendor-overview/, last accessed on August, 2019.

[63] U.S. Environment Protection Agency, “AirNow Current AQI,” https://www.airnow.gov/, last accessed on August, 2019.

[64] EEA, “EEA Air Quality Index,” https://airindex.eea.europa.eu/, last accessed on August, 2019.

[65] D. for Environment Food and R. Affairs, “Pollution forecast - Deffra, UK,” https://uk-air.defra.gov.uk/forecasting/, last accessed on August, 2019.

[66] Breezometer, “Lisbon, Portugal Current Air Quality Map and Air Pollution Forecast,” https://breezometer.com/air-quality-map/city/lisbon, last accessed on August, 2019.

[67] ——, “Most Accurate Real Time Air Quality Data Reporting,” https://blog.breezometer.com/most-accurate-real-time-air-quality-data-reporting, last accessed on August, 2019.

[68] U-blox, “SARA-R4/N4 series Data Sheet,” https://www.u-blox.com/sites/default/files/SARA-R4-N4_DataSheet_{%}28UBX-16024152{%}29.pdf, last accessed on August, 2019.

[69] ——, *SARA-R4 / N4 series - AT Commands Manual*, 2018.

[70] T. Zheng, M. H. Bergin, R. Sutaria, S. N. Tripathi, R. Caldow, and D. E. Carlson, “Gaussian Process regression model for dynamically calibrating a wireless low-cost particulate matter sensor network in Delhi,” *Atmospheric Measurement Techniques Discussions*, no. March, pp. 1–28, 2019.

[71] B. Ford, J. R. Pierce, S. Jathar, M. Long, and E. Wendt, “A low-cost monitor for measurement of fine particulate matter and aerosol optical depth. Part 2: Citizen science pilot campaign in northern Colorado,” *Atmospheric Measurement Techniques Discussions*, no. April, apr 2019.

[72] D. Liu, Q. Zhang, J. Jiang, and D. R. Chen, “Performance calibration of low-cost and portable particular matter (PM) sensors,” *Journal of Aerosol Science*, vol. 112, pp. 1–10, 2017.

[73] PyKrige, “Pykrige - pypi,” https://pypi.org/project/PyKrige/, last accessed on October, 2019.

[74] Mapbox, “Mapbox,” https://www.mapbox.com, last accessed on October, 2019.

[75] Pingdom, “Pingdom Website Speed Test,” https://tools.pingdom.com/.

[76] R. Jayaratne, X. Liu, P. Thai, M. Dunbabin, and L. 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*, vol. 11, no. 8, pp. 4883–4890, 2018.

[77] X. Li, T. R. Dallmann, A. A. May, C. O. Stanier, A. P. Grieshop, E. M. Lipsky, A. L. Robinson, and A. A. Presto, “Size distribution of vehicle emitted primary particles measured in a traffic tunnel,” *Atmospheric Environment*, vol. 191, no. June, pp. 9–18, 2018.