

# Air Quality Monitoring and Interpolation System for IoT Applications

Bernardo Nuno Bernardo Amaral  
bernardo.amaral@ist.utl.pt

Instituto Superior Técnico, Lisboa, Portugal

December 2019

## 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

## 1. Introduction

In the present days, air pollution is a major environmental problem. In 2016, it was responsible for approximately 4.2 million deaths and is currently estimated to be accountable for 25% of adult deaths from strokes, 43% from chronic obstructive pulmonary disease and 29% from lung cancer [20]. It is also responsible for stunting plant growth, lowering agricultural productivity and for reducing city growth and attractiveness to citizens, slowing their evolution and development [4].

At the time of this work, systems for the measurement of air pollution are placed in urban environments as an investment made in the last two decades addressing the emerging knowledge on air pollution health hazards. These systems are constituted by container-style monitoring stations, which occupy large spaces, have high power consumption and high maintenance and production costs. In the greater Lisbon area, there are only fourteen air pollution monitoring stations. From these, only thirteen measure particulate matter with a diameter smaller than 10 micrometers (PM10) and only four measure particulate matter with a diameter smaller than 2.5 micrometers (PM2.5). The monitoring station's network completed 72 stations in 2005, in the overall area of Portugal [1].

The goal of this work is to develop a system to improve the resolution of the visualization of PM10 concentration levels throughout the greater area of Lisbon, with the use of the latest Internet of Things (IoT) technologies available, machine learning and a web application. First, a low-cost portable IoT particulate matter (PM) monitoring system will be assembled. Second, with the help of the Lisbon air quality measures and data from Portuguese Online Database on Air Quality (QualAr) several Spatial Interpolation Models (SIM) will be assessed and compared with Fuzzy Boolean Nets (FBN) in what regards spatial interpolation performance. Finally, a web application that presents live, fine resolution,

interpolated data of PM10 concentration in the city of Lisbon will be developed.

## 2. Background

### 2.1. PM Monitoring Sensors

Despite their price, low-cost optical PM monitors have shown to be able to characterize PM concentrations with high resolution. Recent studies show that these perform with adequate precision for indoor air monitoring, although with some calibration requirements [11], having also demonstrated overall good correlations with reference monitors [16]. Additionally, it has been concluded that comprehensive understanding of the sensor characteristics is necessary and that one big disadvantage is the non-guaranteed accuracy and the lack of established scientific literature regarding this equipment [6].

Zheng et al., in 2018, tested how environment circumstances affect these sensors performance. Results showed that relative humidity (RH) has a major influence in the sensor readings, while the temperature effects were negligible [21].

### 2.2. Interpolation Algorithms for Spatial Analysis

Spatial interpolation consists of estimating the value of measures of certain properties, at unsampled locations, within the area covered by existing observations of the same properties. Some of the most popular spatial interpolation models (SIM) are inverse distance weighting (IDW), kriging, shape functions, spline, and trend surface [8].

In 2014, Li and Heap made a review of the 25 most commonly applied SIMs [9]. As result, an easy to use decision tree was created for use case dependent method selection. 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, according to the decision tree.

#### 2.2.1 Fuzzy Boolean Nets

FBNs are Boolean networks developed and studied by researchers at Instituto Superior Técnico [17]. In these networks, each neuron has a binary value and is constituted by a binary table with a user-defined length.

The networks are constituted by several neuronal areas. Each area has the same number of neurons and can only be one of two types, antecedent or consequent.

Connections between neurons are weightless, temporary and only appear between neurons located in different areas of different types. In Figure 1, a simple example of an FBN is presented, with

two antecedent areas and one consequent area, each with four neurons.

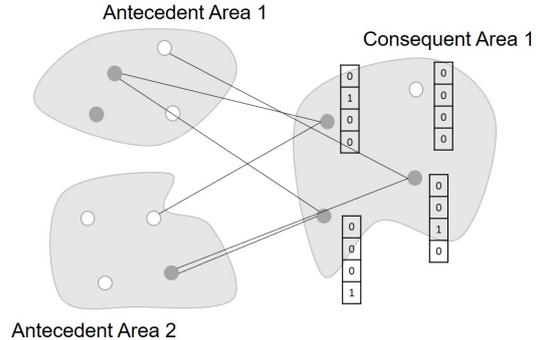


Figure 1: Example of the constituents of an FBN.

The values of training and testing variables are attributed as an activation ratio to each area.

The inference process results of learning and testing epochs where consequent neurons sample an user-defined number of antecedent neurons, and according to these samples, fill their binary tables, in the training process, or retrieve a value from it, in the testing process. The final activation ratio of consequent areas represent the inference results.

In previous studies, FBNs have shown to have positive interpolation capabilities when tested with sparse datasets [18].

#### 2.2.2 IDW

IDW is a deterministic SIM based only on the assumption that the degree of influence of nearby sample points should be greater than the effect of farthest points. The only parameter which is user-defined in the IDW implementation is the power function, which defines the relevance of the distance in the calculation of the weights of each observed value [12].

The IDW interpolation can be defined by the equation (1), where (2) represents the weights associated to each observation.

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

$$w(d_i) = \frac{1}{d_i^p} \quad (2)$$

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.

### 2.2.3 Ordinary Kriging

Kriging geostatistical schemes are stochastic, local, gradual and exact interpolators. In kriging, weights are based not only in the distance between points but also in the overall spatial arrangement of the observation points [12].

Before its application, an exploratory statistical analysis of data must be made, as well as the modelling of a variogram to represent how semivariance varies with distance.

In Figure 2, 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.

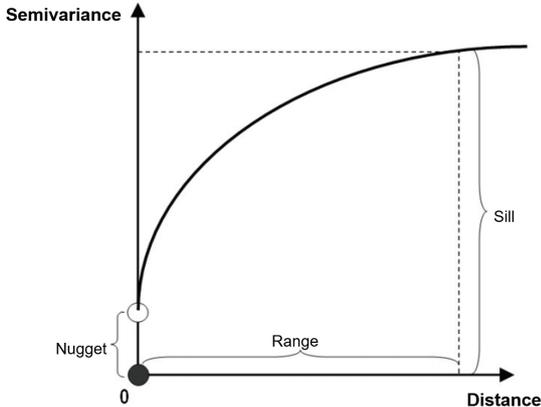


Figure 2: Example of a kriging variogram.

### 2.3. NB-IoT

NB-IoT is a Low Power Wide Area Network (LP-WAN) technology created in 2015 to face the growing massive IoT connectivity challenge. It operates on 180 kHz bandwidth for both downlink and uplink. The downlink remains with the same structure as Long Term Evolution (LTE) (orthogonal frequency-division multiple access (OFDMA) with 15 kHz sub-carrier spacing), and the uplink is single-carrier frequency-division multiple access (SC-FDMA) with sub-carrier spacing at 3.75 kHz (15 kHz for in-band to avoid interference from other LTE traffic) [15].

Studies have compared NB-IoT to other LP-WANs, such as SigFox, Long Term Evolution for Machines (LTE-M) and Long Range Wide Area Network (LORAWAN), and several advantages were observed over the latter. Some of these are

that NB-IoT is more robust in terms of packet error rate [13], it provides the best coverage, even in indoor scenarios [7], and is the most suitable for IoT personal and public applications [3].

### 2.4. Air Pollution Visualization

Currently there are two types of platforms for the visualization of live air quality data: governmental and private companies platforms.

Reviewed governmental platforms do not show fine interpolated data in their air quality live maps [14] [19]. Instead, low resolution maps that do not show air quality variances between stations are presented. On the other hand, business oriented platforms take advantage of the publicly available air quality data provided by governments to make their live interpolated air pollution maps as a products. These represent maps with much finer resolution, but the algorithms used for the interpolation are proprietary [2].

## 3. PM10 Monitoring, Interpolation and Visualization System

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

### 3.1. PM10 Monitoring System Assembly

The low-cost PM10 monitoring system consists of a PM10 sensor, a microcontroller board with a NB-IoT chipset and a step-up voltage converter. The PM10 sensor used was the PMS5003, a low-cost light-scattering optical sensor, which outputs its measures through a digital signal, processed by a microprocessor in real-time. It operates in temperatures between  $-10^{\circ}\text{C}$  and  $60^{\circ}\text{C}$ , and relative humidity from 0 to 99%.

The development board used was SODAQ SFF R412M, which has low power consumption and allows the use of both NB-IoT and LTE-M technologies. It contains an integrated micro-controller, the Atmel SAMD21, which allows it to be programmed through the same tools as Arduino compatible boards.

Subscriber identity module cards provided by the mobile operators NOS and Altice Portugal were used for this board NB-IoT remote communications. These cards 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 1.

Table 1: AT commands used with the NB-IoT micro-controller.

Command	Functionality
AT+CFUN	Set the mobile terminal on
AT+CGATT	GPRS Attach
AT+USOCR	Create a socket
AT+USOST	Write to a remote address
AT+USOCL	Close socket

The resulting system is presented in Figure 3. Two holes were made in the enclosure box to allow the airflow in the inlet and outlet of the sensor.

The development board was programmed, through the Arduino IDE, to collect data from the sensor, calculate 15 minute averages and send them to a server, which stores them in a database, replicating the behaviour of official monitoring networks systems. The server was developed in python, with the usage of UDP sockets, and the used database was MongoDB.

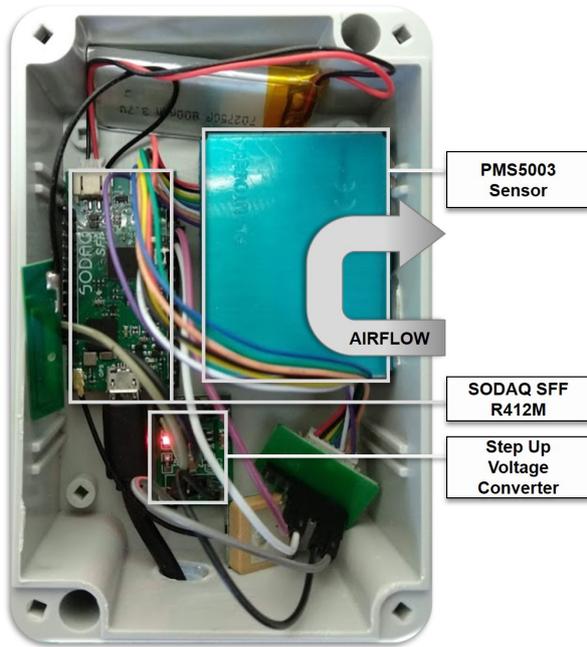


Figure 3: Developed PM monitoring system.

### 3.2. Dataset Selection

Only data from 2013 to 2017 was collected from Portuguese Online Database on Air Quality (QualAr) database, since it is the most consistent, with measures from the considered stations at high percentages of completeness.

Data was filtered to only concern monitoring stations in the greater area of Lisbon, since that is the area where these have the most spatial density in

Portugal. The air quality monitoring network in the greater area of Lisbon is composed of 14 monitoring stations. Only 13 of these have operating PM10 sensors, and only 4 have operating PM2.5 sensors, with several of these lacking registered data in some years. The stations considered in this work are presented in Table 2.

Table 2: Lisbon PM10 stations characterization and completeness in the dataset.

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

Further filtering was applied, data from the stations in Alverca and Alfragide was removed, due to lack of completeness in the dataset and sparse positioning, respectively.

The objective of increasing percentages of data completeness, for every station in the dataset, was accomplished with the results of Table 2. The final dataset contained data regarding 11 657 time intervals.

The geographical location of the stations chosen for algorithm testing can be observed in Figure 4, as well as, in red, the stations which are in the inside of the network outline, which will be used as inference targets.

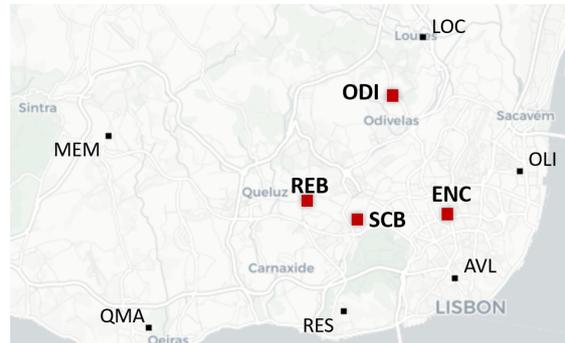


Figure 4: Selected stations, in the greater area of Lisbon.

In the final filtered dataset, the Pearson coefficient was calculated between every station and every inference target station, as one can see in Table 3. Possible resulting values vary from 1 to -1, where

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

Table 3: Pearson correlation coefficients between stations.

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

### 3.3. Algorithm Implementation

Algorithms considered in this work for spatial interpolation were IDW, OK, Linear Interpolation (LI), Nearest Neighbors (NN) and FBNs. In order to assess algorithm performance, cross validation tests were applied to the selected stations. In each iteration, one measure from each central station (Figure 4) was inferred using the readings from remaining stations taken at the same time. In Algorithm 1, the pseudo-code of the setup of this process is presented.

**Algorithm 1:** Algorithm execution setup.

```

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

```

Both LI and NN were implemented with the python *scipy* library *interpolate* package. IDW was programmed in plain code without the use of any packages, due to its operations simplicity.

Ordinary Kriging is the Kriging variant which was considered in this work, and it was implemented with the *pykrige* python library. Kriging variograms can only be modelled if datasets have a relatively high number of points and high spatial density, which is not the case in these works experiments, in which observation points are few and sparsely distributed.

Both IDW and OK were tested beforehand to find the most suitable user-defined parameters for their

application in this work.

#### 3.3.1 Fuzzy Boolean Nets implementation

FBNs have several implementation details which are user-defined and require analysis previous to its application.

##### Antecedent and Consequent Areas

PM10 concentration values and the coordinates (longitude and latitude) are the problem's variables. The coordinates, independent variables, were associated to two antecedent areas and the PM10 concentration, the dependent variable, was associated to one consequent area.

##### Variables Scale Conversion

Since each area only receives inputs between 0 and 1, a scale transformation was calculated for each 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 and the extreme east station is Olivais. Therefore, the longitude defined value for 0 is -9.35 and for 1 is -9.1.

The same was done for the latitude. The southwest station is Quinta do Marquês and the northeast is Loures Centro. Consequently, the defined values for 0 and 1 for the antecedent area corresponding to the latitude coordinate are 38.69 and 38.83 respectively.

Finally, for the PM10 concentration values, the minimum value was  $0 \mu\text{g}/\text{m}^3$ , and the maximum  $175 \mu\text{g}/\text{m}^3$ , which is above the highest registered value in the filtered dataset.

##### Parametrization

Several other user-defined parameters were tested. These were the number of testing and training epochs, the network size, the granularity and the number of samples per neuron.

#### 3.3.2 Algorithm Performance Statistics

In order to assess the results of the cross validation approach, the statistical error parameters presented in equations (3) and (4) were used.

$$MAE = \frac{\sum_{i=1}^n |M_i - O_i|}{n} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (M_i - O_i)^2} \quad (4)$$

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

Maximum Absolute Error (MAE) measures the average magnitude of errors in a set of predictions.

Root Mean Square Error (RMSE) is a quadratic parameter that 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  $\mu\text{g}/\text{m}^3$ . Regarding RMSE, since errors are squared before they are averaged, increased weight is given to large errors.

### 3.4. System for online data visualization

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

#### 3.4.1 Database

A MongoDB database was developed to store the values obtained by the monitoring systems.

MongoDB was chosen for this specific application due to its high scalability and flexibility. 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 is saved, with corresponding date and time, and sensor ID. Despite the non-enforcement of document structures in the database, for the purpose of this work, each valid document must have fields for the date, location, sensor ID and PM10 value, and the set of date and sensor ID fields can not be repeated in different documents.

#### 3.4.2 Mapbox GL JS

In this work, extrusion of polygons, provided by the Mapbox GL JS library, was used to represent the levels of estimated PM10 concentration values. Extrusion height and color is determined by the pollution value at each location of the map.

For the representation of these extrusions, a convex hull was defined, delimited by the Lisbon monitoring station networks which are on the edges. This resulted in a grid with a resolution of  $100 \text{ m}^2$ . 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 performs an hourly interpolation with the updated values, according to QualAr reports, and updates the geojson.

## 4. Results & discussion

For a better understanding of the results of this work it is important to note how PM10 concentra-

tions affect air quality. In Table 4 one can see the official classification of the PM10 concentration values [14].

Table 4: Classification of PM10 concentration.

Classification	PM10 concentration interval ( $\mu\text{g}/\text{m}^3$ )
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  $14 \mu\text{g}/\text{m}^3$ , for both Good and Medium intervals.

### 4.1. NB-IoT PM10 Monitoring System

In Table 5 are presented the weeks during which the developed system was placed in a CCDR-LVT station.

Table 5: Coverage during the developed system placement.

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

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. In red colored weeks, it was impossible to get consistent coverage and consequently no measures were obtained.

In Figure 5, simultaneous measures of both sensors during the placement period can be observed, along with RH measures obtained for Lisbon during the same period. The absolute error between the measures of the different sensors is  $14.40 \mu\text{g}/\text{m}^3$ .

The PMS5003 sensor often obtained lower measures. However, sometimes it peaked, surpassing the reference sensor, the Environnement S.A. Model MP101M measures. This only happened for high values of RH. When the RH decreases, there is often less error between sensors.

The scatter diagram in Figure 6 shows that the developed system measured values are predominantly lower than the reference readings.

In the comparison of time series data, it is important to assess not only the error statistics between datasets, but also its linearity and correlation.

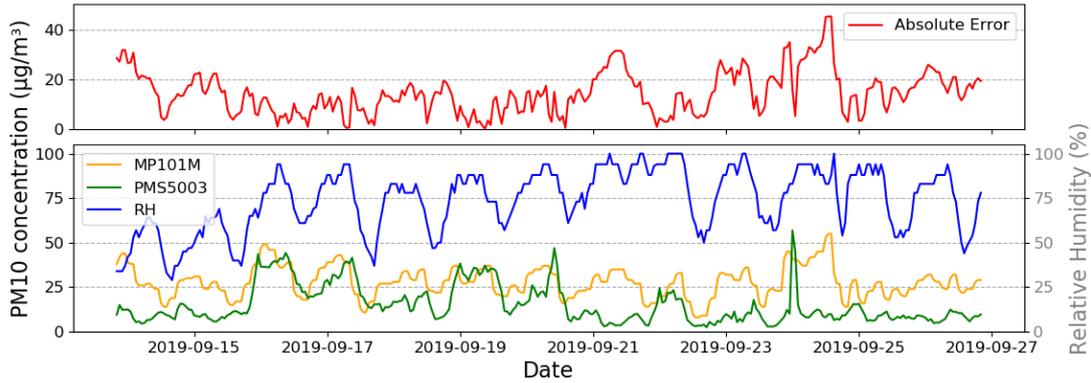


Figure 5: PM10 concentrations reported by the sensors and simultaneous RH values.

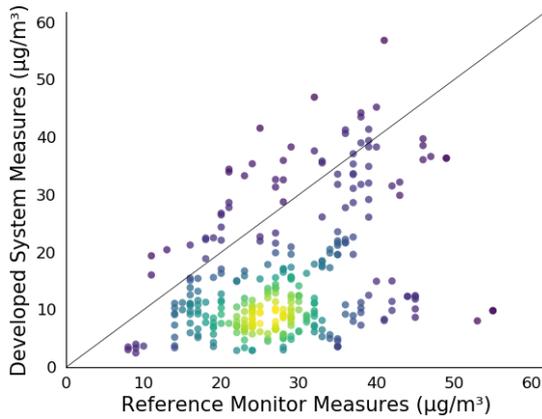


Figure 6: Scatter diagram between the developed system and the reference monitor measures.

The Pearson coefficient was calculated for every full day of the time series data, during the developed sensor placement. The coefficient obtained for the whole period was 0.39, with several days presenting negative correlation coefficients between measures.

#### 4.2. SIMs Testing

In this section, every SIM performance test result is presented.

Through the cross validation approach application in the dataset, the IDW power function was tested with values between 0 and 5, and the value with the best error statistics was  $p = 0$ .

The same approach was applied in the parameter assessment of the remaining algorithms. In OK, the variogram is the main user-defined parameter. Several pre-defined variograms were tested, and in the obtained results the linear model had the lowest overall MAE and RMSE values.

In what regards FBN interpolation, both testing and training epochs, network size, number of samples and granularity of the memories of each consequent neuron were tested. The chosen parameters were 30, 30, 75, 9 and 3, respectively.

#### 4.2.1 SIMs Comparison

Results regarding MAE and RMSE performance are presented in Table 6, for each algorithm and each target station, for a total of 43 956 inferences.

Table 6: 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

The algorithms with overall lower MAE and RMSE results were IDW and OK. NN had the worse results, followed by LI. Histograms resulting from the IDW, OK and FBN execution can be observed in Figure 7.

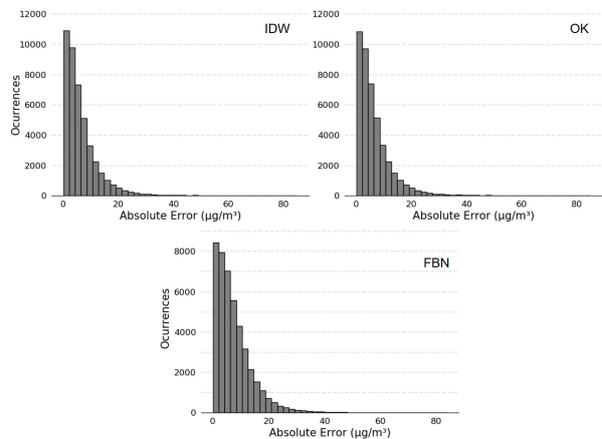


Figure 7: Histograms obtained from 43 956 inferences for IDW, OK and FBN.

It should additionally be noted that, for all the inferences done simultaneously, FBNs execution times were around 9 hours, while remaining algorithms only took approximately 5 minutes.

#### 4.3. Online Data Visualization Web Application

The resulting PM10 visualization application has a resolution of 100 meters, both in color and in height of the pollution surface. It provides the possibility to hover over the grid to consult 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 8.

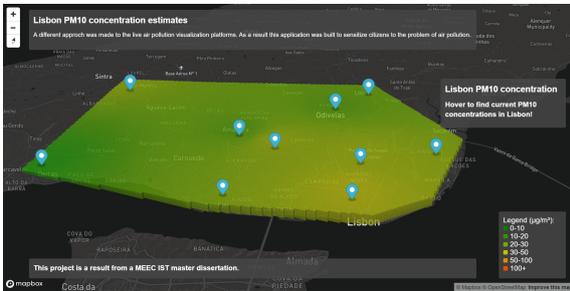


Figure 8: Developed Web Application final result.

#### 4.4. Discussion

##### 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 was observed that it was not good enough for professional applications at the time of this work. During the 10 weeks of testing in different locations in the city of Lisbon, there were only 2 weeks when it was possible to get consistent coverage. According to the literature [7], NB-IoT has one of the widest coverage between IoT technologies. This was not observed in this work. Results demonstrated 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 measurement performance, it is important to note that the obtained data only corresponds to a small sample size in comparison to what was expected.

As mentioned, during the sensor placement and comparison period, the PMS5003 sensor often obtained lower measures (Figure 6). However, sometimes it peaked, surpassing the MP101M measures.

This only happened for high values of RH. It could additionally be observed, in Figure 5, that when the RH is decreasing, there is often less error observed between the sensors.

The average absolute error between the measures of the different sensors ( $14.40 \mu\text{g}/\text{m}^3$ ) is slightly above the smallest scale division of the classification in Table 4. The daily pearson correlation coefficients showed inconsistency in the linearity between sensor readings.

It was summer season during this experiment, and the only rain occurrences during the placement periods happened in the day 21 of September. It was the day with the second highest absolute error between readings and PMS5003 measured values were lower than average in comparison to the reference sensor.

In a study by Zheng et al., in 2018, a similar sensor model (PMS3003) was extensively tested and it was observed that while temperature took a negligible part in the difference observed between sensors, RH had a much more considerable influence [21].

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.

In the same study, it was stated that above 75% RH, 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 [5]. However, in this work, the only observed effects were the readings increase, which was also reported by the MP101M, and the mentioned undervalued PMS5003 measurements during the rain periods.

No substantial difference was observed between week and weekend days, despite the average PM10 concentration being slightly lower in both sensors.

During the day, MP101M readings are mostly higher than the taken by PMS5003. According to Jayaratne et al., in the same study [5], this may happen because the size of traffic particles are below the minimum detectable size limit of the PMS sensors. It is documented that the diameter of particles emitted by vehicles can be mostly below 300 nm [10].

Due to the numerous disadvantages observed with data from this work, it can be inferred that low-cost sensors are not suited for regulatory applications, such as assessing if the air quality meets stipulated guidelines. However, with previous study of the placement location and the deduction of a fitting correction factor to each case, it could be implemented for informative purposes, such as providing a better resolution to monitoring networks which suffer from sparsity.

#### 4.4.2 SIMs Assessment

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

In Table 3, correlation coefficients between every station of the dataset were calculated. It can be observed that the stations with shared highest correlation coefficients were relatively close in terms of geographical location, despite an overall high correlation between almost every station. However, the good performance of IDW with  $p = 0$ , which means that distance between data points is not considered for the interpolation, suggests that the distance correlation between stations is very low.

Air dispersion and movement in urban environments is an environmental problem of high complexity. In the last 5 years, in the city of Lisbon, the dataset reveals that the city of Lisbon has an overall low concentration of PM10, with a maximum mean value for the whole dataset of  $34.22 \mu\text{g}/\text{m}^3$  in the station of Avenida da Liberdade, 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, as is the IDW interpolation.

Several other environmental factors can have influenced the correlation of PM10 concentration 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 geographical and geographical barriers to airflow.

In what regards FBNs performance, no conclusions can be obtained towards its performance regarding multidimensional 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 observed high computing times.

#### 4.4.3 Web Visualization Application

The developed application resulted in a platform with the presentation of live PM10 concentration data with high resolution in the area covered by the CCDD-LVT monitoring stations network in Lisbon. Since results obtained from the low-cost sensors reveal these are not yet suitable to be integrated on monitoring networks, these were not included in the network used for the spatial interpolation presented in the application.

The application 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 be applied in other geographical interpolation fields.

In comparison with current governmental web applications available, the developed application has higher resolution and provides better user experience, despite not covering large areas such as whole countries. As a product, there are already companies with applications which outperform it, being the most evident Breezometer. It presents data regarding numerous air polluting gases and indicators with a resolution of 500 meters, and worldwide coverage. However, due to its proprietary characteristics, its air modelling algorithms and tools are not documented and consequently not available for assessment.

## 5. Conclusions

The current state of the data available publicly regarding air quality was studied and assessed. Several experiments were made in order to evaluate the application of new technologies which could improve it. The tested PMS5003 low-cost PM sensor did not show promising results for official networks integration, and requires complex calibration in order to be used for interpolation purposes.

Spatial interpolation algorithms showed a low distance correlation, due to the sparsity of the stations in the city, adding to the importance of the study of lower-cost sensors integration.

It can be concluded that the developed visualization 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.

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.

### 5.1. 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. Low-cost PM monitoring sensors should be further studied in controlled environments. Ideally, with 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 network for interpolation purposes.

Spatial algorithms tested in this work could be complemented with the addition of other indepen-

dent variables, besides coordinates, and air dispersion models. FBN implementation could be parallelized, and further studied regarding spatial geographical interpolation with the integration of additional variables to the problem.

## References

- [1] APA. *Evolução da qualidade do ar em Portugal entre 2001 e 2005*. 2008.
- [2] Breezometer. Lisbon, Portugal Current Air Quality Map and Air Pollution Forecast. <https://breezometer.com/air-quality-map/city/lisbon>, last accessed on August, 2019.
- [3] 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*, 23:60–67, 2016.
- [4] GSMA. Air Quality Monitoring Using IoT and Big Data. [https://www.gsma.com/iot/wp-content/uploads/2018/02/iot\\_clean\\_air\\_02\\_18.pdf](https://www.gsma.com/iot/wp-content/uploads/2018/02/iot_clean_air_02_18.pdf), last accessed on August, 2019.
- [5] 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*, 11(8):4883–4890, 2018.
- [6] J. Kuula et al. Applicability of Optical and Diffusion Charging-Based Particulate Matter Sensors to Urban Air Quality Measurements. *Aerosol and Air Quality Research*, 19(5):1024–1039, 2019.
- [7] M. Lauridsen, H. Nguyen, B. Vejlggaard, 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*, 2017-June:2–6, 2017.
- [8] J. Li. A review of spatial interpolation methods for environmental scientists. *Canberra: Geoscience Australia*, 137, 2008.
- [9] J. Li and A. D. Heap. Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling and Software*, 53:173–189, 2014.
- [10] 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*, 191(June):9–18, 2018.
- [11] A. Manikonda, N. Zíková, P. K. Hopke, and A. R. Ferro. Laboratory assessment of low-cost PM monitors. *Journal of Aerosol Science*, 102:29–40, 2016.
- [12] S. M. P. Mesquita. *Modelação da distribuição espacial da qualidade do ar em lisboa usando sistemas de informação geográfica*. PhD thesis, Instituto Superior de Estatística e Gestão de Informação da Universidade Nova de Lisboa, 2009.
- [13] H. Mroue et al. MAC layer-based evaluation of IoT technologies: LoRa, SigFox and NB-IoT. *2018 IEEE Middle East and North Africa Communications Conference*, pages 1–5, 2018.
- [14] QualAr. Mapa dos índices diários provisórios (não validados). <https://www.qualar1.apambiente.pt/qualar/>, last accessed on August, 2019.
- [15] 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.
- [16] T. Sayahi, A. Butterfield, and K. Kelly. Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environ Pollut*, 2018.
- [17] J. Tomé. Neural Activation Ratio Based Fuzzy Reasoning. *1998 IEEE International Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence*, pages 1217–1222, 1998.
- [18] J. A. Tomé and J. P. Carvalho. Fuzzy Boolean Nets - A nature inspired model for learning and reasoning. *Fuzzy Sets and Systems*, 253:1–27, 2014.
- [19] U.S. Environment Protection Agency. AirNow Current AQI. <https://www.airnow.gov/>, last accessed on August, 2019.
- [20] WHO. 9 out of 10 people worldwide breathe polluted air, but more countries are taking action. <https://www.who.int/news-room/detail/>, last accessed on August, 2019.
- [21] T. Zheng et al. Field evaluation of low-cost particulate matter sensors in high-and low-concentration environments. *Atmospheric Measurement Techniques*, 11(8):4823–4846, 2018.