Masters Program in **Geospatial Technologies**

Improving public health in smart cities in the air pollution context

Anthony Twesigye

Dissertation submitted in partial fulfilment of the requirements for the Degree of *Master of Science in Geospatial Technologies*

NOVA IMS Information Management School

Universitat Jaume I

Westfälische Wilhelms-Universität Münster

IMPROVING PUBLIC HEALTH IN SMART CITIES IN THE AIR POLLUTION CONTEXT

Dissertation supervised by

Prof. Ramos Francisco, Universitat Jaume I, Castellón, Spain

co-supervised by

Prof. Cristina Costa, Universidade Nova de Lisboa, Portugal

Prof. Judith Verstegen, University of Münster, Germany

February 2018

Acknowledgements

I thank the Almighty God for the gifts of life, prayer, protection and continued guidance in my life that has always been my source of encouragement to take the next step in life.

I thank all my supervisors Prof. Ramos Francisco, Prof. Cristina Costa and Prof. Judith Verstegen for their continued guidance that gave me renewed energies at every step throughout the master thesis. Special thanks go to Prof. Ramos Francisco, Dr Sergi Trilles Oliver whom I worked closely with from the start of the master thesis till the end and Andres Munoz during the routing service implementation.

I thank my family members and friends whose continued love, support, encouragement and confidence in me has increased my efforts in becoming a better person both in my career and in my personal life.

I thank all my professors during this master program for helping me acquire knowledge and skills that will help me take the next step in my career in offering better and advanced solutions in diverse fields.

I thank my colleagues in the master program for the cooperation during our studies.

Abstract

The public has continually developed interest in knowing the air quality around them. This is of great importance not only for planning their activities, but also for taking precautionary measures for their health. With support from smart cities infrastructure that supports taking measurements of pollutant concentrations, several countries and researchers have used the concept of air quality index (AQI) in its different forms of air quality or air pollution to interpret and communicate such measurements.

In this study we have reviewed the implemented indices by government bodies and some formulations from researchers in relation to the available data to determine an optimum index for Madrid city. This comparison has helped to formulate the Madrid Local Air Quality Index (MLAQI), which considers the local situation in Madrid city.

In relation to the available data from the city council, we have reviewed and compared some of the spatial interpolation methods that have been applied in the field of air pollution. This helped us to identify IDW for support of automated hourly pollution interpolation for the available data from Madrid pollution sensors.

We have then used MLAQI and IDW to create an hourly pollution Web Feature service aimed at helping with public awareness of the air quality around them. The surfaces are categorised with the index categories from good to very poor categories with defined colour coding.

We used the created service to develop a routing web application where high MLAQI categories of poor and very poor are used as polygon barriers to limit the route calculation in those polluted areas thereby helping the public to protect their health from such areas.

Key words

Public health

Air Quality Index

Air Pollution Index

Pollutant concentrations

Spatial interpolation

Web feature service

Acronyms

AQI – Air Quality Index

PSI – Pollution Quality Index

MLAQI – Madrid Local Air Quality Index

CO – Carbon Monoxide

SO_2 – Sulfur Dioxide

O_3 – Ozone

PM10 – Particulate Matter 10 micrometers or less in diameter

PM2.5 – Particulate Matter 2.5 micrometers or less in diameter

NO_2 – Nitrogen Dioxide

WFS – Web Feature Service

REST – Representational State Transfer

JSON – JavaScript Object Notation

ESRI – Environmental Systems Research Institute

URL – Uniform Resource Locator

Table of Contents

Index of tables

Index of figures

1 Introduction

1.1 Background to the study

Smart cities use information technologies to improve on the performance and quality of urban services, to decrease costs and optimize resources, and more so they involve citizens to participate actively in such activities(Zanella *et al.*, 2014). Among the various areas developed in such a context, this study focuses on public health. There has been increasing public need for better air quality or for avoidance of air pollution, which has led to the establishment of determination and representation approaches of these occurrences continuously using the index concept in their formulations.

In recent years, air pollution control has demonstrated to have a positive impact on public health (Correia *et al.*, 2013). A control measure taken by governments or local administrations involves using specific sensors distributed over a wide area usually named air pollution sensors, which are able to detect different levels of air pollution in a particular location.

Deployment of these sensors is slightly growing up with Smart cities and Internet Of Things(Zanella *et al.*, 2014), which enables us to obtain access to their produced data. However, the information usually provided to the users is one-dimensional based, in this case corresponding to a determined and fixed latitude and longitude where the sensor is installed.

Many works have been done to publish or generate two-dimensional data from those types of sensors, among these we underline the ones using interpolation methods that use spatial analysis applying statistical theory and techniques to model spatially referenced data.

In our context, pollutant concentrations are types of data that can be represented by surfaces where each raster cell represents a measurement such as a cell's relationship to a fixed point or specific concentration level. Due to impracticability of obtaining values for each cell in a raster, sample points are used to derive the intervening values using interpolation methods. This ability to create surfaces from sample data of air pollution sensors makes spatial interpolation both powerful and useful for the study.

In disseminating pollution information, several government bodies and industry players publish pollutant concentration levels on their websites or mobile applications. The published information can be in the form of pollutant concentrations or scaled concentrations based on a particular air quality or air pollution index. Some of these indices provide health related recommendations to the general public or specific groups of people for the different levels of pollutant concentrations.

1.2 The study area

The chosen study area is Madrid city, the capital of Spain. The city has had challenges with air pollution and has continuously set up measures to control this pollution challenge. One of the measures used for management and control of this challenge is the use of a network of automatic monitoring stations and other mobile stations for measuring pollutant concentrations and calibration of the measurement equipment.

In engaging the public on the awareness of the air quality around them, the city council implemented an air quality information system which disseminates such information to the public in various ways like sms, emails, website, “Aire de Madrid” mobile application and public displays. The hourly pollutant concentration levels are communicated to the public through the hourly index while the daily levels are communicated through the daily index. The daily communication includes maps of the air quality prediction surfaces for a particular day and the following day are categorised by zone in Madrid. The city council also publishes pollutant concentration values through the open data portal for researchers and application developers to use in studies about this challenge in the city.

1.3 Problem statement

The public has continuously developed interest in knowing the state of the air quality, which has supported the development of various approaches for air pollution concentration measurements, representation and dissemination of the resulting information to the public.

As the Madrid city council offers the necessary infrastructure and information services to take pollution concentration measurements, this has enabled reporting about the state of air quality in the city to the public. The reports are mainly the pollutant concentrations at the locations of sensor stations scaled on a given index with a colour representation and the daily prediction pollution surfaces together with a prediction surface for the following day categorised by the different zones of Madrid.

With the city’s pollutant concentration values becoming too high at certain hours, daily pollutant prediction surfaces categorised by zones may not effectively help the public to plan their hourly activities or avoid certain areas with high pollutant concentration values at certain hours. This presents a need for an hourly pollution surface service which service can be used by the general public in planning their hourly activities or be incorporated in other services of public interest like routing for the public to use in their navigation between locations of interest.

1.4 Aim

To develop an hourly air pollution surface service for Madrid city that will help the public to be aware of the air quality around them and be in position to make informed decisions while planning their activities around the city.

1.5 Objectives

In achieving the study's aim, we will meet several objectives.

- Review and analysis of existing index approaches used in communicating air quality or air pollution.
- Review and analysis of different interpolation techniques, both deterministic and geostatistical ones used in the field of air pollution.
- Modification of the existing Madrid Air Quality Index for better air pollution representation with support of spatial interpolation.
- Obtaining and aggregating the Madrid city council open portal published pollution data to support index calculation using the modified index.
- Perform spatial interpolation on the computed index and creating hourly raster surfaces for the index.
- To support the dissemination of these results to smart citizens and improve their health, there is a need for creating a real-time conversion service to generate vector geometries from the interpolated raster surfaces into categories of Good, Acceptable, Poor and Very Poor according to the index.
- To facilitate the data access and exploitation from final applications, we had to create an automatic hourly publishing service for publishing and sharing the created vector surfaces online.
- Implementing a web application using the published service to help the public plan a path to walk or run by minimizing the high pollution areas to traverse, a way to improve their health.

1.6 Hypothesis

Air pollution concentration measurements from a Smart city's infrastructure can support the generation of hourly pollution surfaces, which can help protect the health of citizens. These surfaces can be incorporated into activities of public interest to help the public make informed decisions about such activities.

1.7 Limitations

The study is limited to the available published data from active sensor stations in Madrid city. The number of stations may affect the accuracy of spatial interpolation.

The access of the ESRI routing and traffic premium services is limited to use with a proxy web server. The proxy setup requires an active ESRI developer account for generating required tokens and application registration access information. This routing service also limits the number of intersected street with polygon barriers in routing.

2 Literature review

2.1 Comparative study of Air Quality Indices

In pursuit of addressing air pollution issues around the globe, several approaches for reporting the studies and the results have been developed and tested. One of the approaches developed is the approach of air quality index that seeks to represent the level of the air quality in a location of interest. The indices developed consider different pollutants and use varying limits in reporting the results.

The study compares some of the formulated indices from governmental bodies and the research community, some of which several government bodies have implemented. To compare these approaches, the study is based on their definitions and calculation, categories considered with the category ranges, the symbology used in their representations, the general health recommendations to the public and to specific groups of people, the effect of multi-pollutants and concentration measurement location variations.

2.1.1 The United States Environmental Protection Agency (EPA) Air Quality Index (AQI)

The US EPA AQI categorises air quality in six categories of Good with range (0-50), Moderate with range (51-100), Unhealthy for Sensitive Groups with range (101-150), Unhealthy with range (151-200), Very Unhealthy with range (201-300) and hazardous with range (301-500). With this AQI, the values above 500 are also considered hazardous (US EPA, 2014). These categories increase with increasing effect on human health and are assigned standard colours for easier identification and reporting (US EPA, 2013).

This AQI is defined for pollutants of Ozone, PM2.5, PM10, Carbon monoxide, Nitrogen Dioxide and Sulfur Dioxide. These pollutants (O_3, PM1, PM2.5, PM10, CO, NO_2, and SO_2) are also critical for research and Industrial IoT system deployment with US EPA-funded testing facilities like AQ-SPEC at the Southern California Air Quality Management District (SC - AQMD) in Los Angeles, California (Valarm, 2018). The EPA also defines the limit values for specific time scales of these pollutants for computation of the AQI. For Ozone, the limit values are defined for 1hour and 8hours, for PM 2.5 and PM10, the limit values are defined for 24hours, for Carbon monoxide, the limit values are defined for 8hours and the limits for Nitrogen Dioxide and Sulfur Dioxide are defined for 1hour. Higher values of limit values do not indicate higher AQI but are the basis for calculation of the index (US EPA, 2013). At the established categories, the AQI defines health related risks or groups of people that are highly affected by the levels of air pollution.

2.1.2 The Canada Air Quality Health Index.

The Canadian government uses an Air Quality Health Index (AQHI) developed based on the relative risk of pollutants to human health with a scale designed to help people understand what the air quality around them means to their health. It helps people make decisions regarding short-term exposure to air pollution and adjusting their activities based on the information obtained (Stieb *et al.*, 2008; *ECCC Canada*, 2016).

The AQHI communicates the air quality health related risks on a scale of 1 to 10+ with 4 categories of health risks as Low Health Risk (1-3), Moderate Health Risk (4-6), High Health Risk (7-10), or Very High Health Risk (10+). A colour scheme of light blue for lower values of the index to brown for higher values of the index is used in the index communication. It also gives the health messages for both the population at risk and the general population. AQHI uses the relative risks of a combination of pollutants of Ozone, PM2.5 and Nitrogen Dioxide to determine the final index.

2.1.3 Common Air Quality Index (CAQI)

Developed under the CITEAIR project of the European Union with the aim of establishing a platform for comparing air quality across the different cities in Europe, a review of the existing indices was carried out and CAQI developed to achieve the comparability across the cities in real-time and caters for hourly, daily and yearly time scales. It established two indices, one for the roadside monitoring stations and the other one for city background conditions (van den Elshout, Léger and Nussio, 2008). CAQI defines five classes with appropriate ranges of Very Low (0-25), Low (25-50), Medium (50-75), High (75-100) and Very High (above 100).

The first version of CAQI considered pollutants of Ozone (O_3), PM10, Carbon monoxide (CO), Nitrogen Dioxide (NO_2) and Sulfur Dioxide (SO_2) but the revised version of the index introduced PM2.5 in the index calculation (van den Elshout, Léger and Nussio, 2008; Van Den Elshout, Léger and Heich, 2014). CAQI uses the concept of core pollutants for the two indices, introduced to allow for the calculation of the index and without it an index calculation is not performed. For the roadside index, the core pollutants considered are NO_2 and PM10 with CO and PM2.5 as auxiliary pollutants while for the city background index NO_2, PM10 and O_3 are considered core pollutants with CO, SO_2 and PM2.5 considered as auxiliary pollutants. The index is calculated by linear interpolation between the class borders of the pollutants and the final index given as the highest sub index of the considered pollutants (van den Elshout, Léger and Nussio, 2008; Plaia and Ruggieri, 2011).

2.1.4 The UK Air Quality Index

In a review of the UK Air Quality Index, Committee on the Medical Effects of Air Pollutants (COMEAP) developed and recommended a Daily Air Quality Index (DAQI) for the purpose of providing short-term health advise to the public regarding the air quality around them and possible recommendations such people can take (COMEAP UK, 2011). Following the recommendations from COMEAP, the Department for Environment, Food and Rural Affairs (Defra) together with responsible administrations implemented this index from January 2012. This Index was later updated with minor changes through Defra's update on the implementation of DAQI in April 2013 to conform with the EU limit values of pollutant concentrations. The update also emphasised that data rounding off should always be performed at the end of calculations before communicating results to avoid errors (Emily Connolly *et al.*, 2013).

DAQI is defined on a scale of 1 to 10 with colour coding and categorised into four bands of Low (1-3), Moderate (4-6), High (7-9) and Very High (10) (COMEAP UK, 2011). COMEAP report recommended the removal of CO from the AQI and the inclusion of PM2.5. Currently the DAQI uses the pollutants of Ozone (O_3), Nitrogen Dioxide (NO_2), Sulfur Dioxide (SO_2), PM2.5 and PM10 in calculation of the index. The Automatic Urban and Rural Network (AURN) measures these pollutant concentrations in near real-time and their values used in calculation of the index (COMEAP UK, 2011). The overall index is given by the highest pollutant concentration of the considered pollutants. To allow prediction of elevated air pollution episodes in real-time, DAQI uses trigger values to predict concentrations of pollutants (COMEAP UK, 2011).

2.1.5 Ireland Air Quality Index

In order to find an appropriate Air Quality Index for Ireland that included health information, Air Quality Health Information Working Group was set up in 2011 for the task. The Health Service Executive (HSE) reviewed existing health based evidence on the impact of air pollution on health and also reviewed a selection of Air Quality Indices that existed across the globe. The committee reached similar conclusions to those done by the COMEAP. This led to the proposal by the Irish Environmental Protection Agency (EPA) for adopting the UK DAQI into the Irish air quality monitoring infrastructure. The HSE and Irish EPA reached the conclusion of introducing the Air Quality Index for Health (AQIH) that is closely aligned with the UK DAQI.

The AQIH was proposed to maintain consistency with the previous Irish index that was also an air quality index. It is noted that the UK DAQI is a pollution index as it was replacing a previous air pollution index. This difference led to the difference in the naming of the index categories (EPA

Ireland, 2013a). The Irish AQIH uses a scale with defined colour coding of 1 to 10 that is categorised into four bands of Good (1-3), Fair (4-6), Poor (7-9) and very poor (10) (*EPA Ireland*, 2013b). These bands correspond to the UK DAQI bands of Low, Moderate, High and Very High respectively. The other information like health messages and interpretation of the index is the same. Like the DAQI, AQIH considers five pollutants of ozone, nitrogen dioxide, sulphur dioxide, PM2.5 and PM10. The final index is given by the worst index of the separately calculated indices of the considered pollutant concentrations.

2.1.6 Spain, Madrid Air Quality Index

Madrid City Council provides information to its community using an air quality index defined by a scale of 0 to >150 with four categories of “Buena” - good (0-50), “Admisible” - acceptable (51-100), “Deficiente” - poor (101-150) and “Mala” - very poor (>150). The City council uses the pollutants of PM10, Sulfur Dioxide, Nitrogen Dioxide, Carbon Monoxide and Ozone. Sub indices are calculated for the considered pollutants and the final index is the worst sub index of the pollutant concentrations (*Madrid City Council*, 2015).

2.1.7 France Air Quality Index

The ATMO Index is the air quality index used in major cities in France that have a population of more than 100,000 inhabitants. The index is represented by a giraffe and is based on a scale of 1 to 10 ranging from very good to very bad and with three coloured bands of Green (1-4), Orange (5-7) and Red (8-10). ATMO Index considers the pollutants of Sulfur Dioxide, Nitrogen Dioxide, Ozone, PM2.5 and PM10. Sub indices are calculated for the four pollutant concentrations and the final aggregated index is the highest sub index calculated from the pollutant concentrations (*ATMO France*, 2008). Each pollutant has defined limit values for the scale ranges upon which pollutant concentrations are compared to determine the pollutant sub index.

2.1.8 Singapore Air Quality Index

Singapore reports air quality in terms of Pollutant Standards Index (PSI). The index is categorised into five categories of Good (0-50), Moderate (51-100), Unhealthy (101-200), Very Unhealthy (201-300) and Hazardous (301-500) (*NEA Singapore*, 2017). PSI is based on six pollutants of PM2.5, PM10, Sulfur Dioxide, Carbon Monoxide, Ozone and Nitrogen Dioxide. The sub indices of all the considered pollutants are calculated and the final index is the highest sub index of the pollutant concentrations (*NEA Singapore*, 2014).

2.1.9 Researchers work on Air Quality Index representation

Several researchers have continued to investigate on the better representation of air quality index. We have specifically selected five published papers detailing some of these methods which are related with health, the effect of multi pollutants on heath and spatial variability of pollutant measurement locations.

To account for multi pollutant short-term health effects of exposures on the final index, (Cairncross, John and Zunckel, 2007) formulated an Air Pollution Index (API). To account for this effect, the final index is the summation of the normalised sub indices of pollutant concentrations. The developed index depends on the relative risk of daily mortality associated with the common pollutants of PM2.5, PM10, Sulfur Dioxide, Nitrogen Dioxide, Ozone and Carbon Monoxide. The proposed index has a scale of 1 to 10 with defined colour codes. It also defines the index with four categories together with associated increase in mortality risks. The four categories are low (1-3), moderate (4-6), high (7-9) and very high (10).

Based on time series analysis of air pollution and mortality in Canadian cities, (Stieb *et al.*, 2008) proposed an Air Quality Health Index (AQHI). To cater for multi pollutant effects and varying seasons, they carried out analysis for pollutant combinations using single and multi pollutant models and for varying seasons. To develop the index, they used the combination of Carbon Monoxide, Nitrogen Dioxide, Ozone, PM2.5 or PM10 and Sulfur Dioxide pollutants. The index excludes Sulfur Dioxide and Carbon Monoxide in its formulation after realising their small effect during the analysis. The index is defined by four different scenarios using PM2.5, PM10, warm and cool seasons for the case of PM2.5. The index created is on a scale of 0 to 10+ with categories of Low risk (0-3), Moderate risk (4-6), High risk (7-10) and Very high risk (above 10). A corresponding colour scheme is also defined along the scale with colours ranging from light blue at low AQHI values to brown at high AQHI values and red for very high risk category. It provides health related messages to the population at risk and the general population.

Using the pollutants of Carbon Monoxide, Sulfur Dioxide, Nitrogen Dioxide, Ozone and PM10, (Kyrkilis, Chaloulakou and Kassomenos, 2007) developed the aggregate Air Quality Index for Athens, Greece. To cater for multi pollutant effects, they adopted an aggregate function to compute the overall index of the city. They compared their results with the modified USEPA AQI using the European pollutant standard limits and found the modified USEPA predicted higher values than the developed aggregate AQI.

In studying multi-pollutant effects and relative risks of short-term exposure to pollutants, (Sicard *et al.*, 2011) developed the Aggregate Risk Index. The ARI is based on the exposure response relationship and relative risk of established effects to assess the additive effects of pollutants. The method used published relative risk functions data and particular sets of relative risks for associated health risk end points to derive the index. The ARI considers the relative risks of Sulfur Dioxide, Nitrogen Dioxide, PM2.5, PM10 and Ozone pollutants. In catering for the multi pollutant effects, the final index is the summation of individual calculated risk indices. The index is defined from 0 to 10 with the risk values used to derive the break points. The index is categorised into Low (1-3), Moderate (4-6), High (7-9) and Very high (10) with appropriate information about the excess relative risk of mortality or morbidity.

In considering the additive effects resulting from multi pollutants and the effect of measuring pollutants over different geographical locations, (Murena, 2004) developed a daily Air Pollution Index (PI) modified out of the US EPA AQI. The developed index uses European limit values in its computation. The index uses the common pollutants of Carbon Monoxide, Nitrogen Dioxide, PM10, Sulphur Dioxide and Ozone in its computation. The index is defined on a scale of 0 to 100 and it defines five categories of Good quality (25), Low pollution (50), Moderate pollution (70), Unhealthy for sensitive groups (85) and Unhealthy (100). The index introduced clouds for representing the pollution categories. The method introduced considers the sum of ratios of daily reference concentrations of pollutants and their bottom breakpoint concentration values to cater for multi pollutant additive effects on human health and introduces weights for geographical location variability of sensor measured pollutants concentrations.

2.1.10 Summary of the reviewed indices

The reviewed indices share and differ in some of their formulations and representations Table 1. To relate these indices, we have summarised them in terms pollutants considered in their formulation, the number of categories and ranges they consider, the symbolisation and graphical representation used, the health recommendations for the categories, the effect of multi pollutants and the spatial variability of concentration measurement locations.

Index	Pollutants Considered						Number of Categories	Ranges of Categories	Symbolisation	General Health Recommendations	Specific Groups Recommendations	Multi-Pollutant Consideration	Measurement location variation
US EPA, AQI	CO	NO2	O_3	PM2.5	PM10	SO_2	6	0 - 500	Colours	Yes	Yes	No	No
Canada, AQHI		NO2	O_3	PM2.5			4	1 - 10+	Colours	Yes	Yes	Yes	No
Common Air Quality Index, CAQI	CO	NO2	O_3	PM2.5	PM10	SO_2	5	0 - 100+	Colours	No	No	No	Yes
UK Defra, DAQI		NO_2	O_3	PM2.5	PM10	SO_2	4	1 - 10	Colours	Yes	Yes	No	No
Irish EPA , AQIH		NO_2	O_3	PM2.5	PM10	SO_2	4	1 - 10	Colours	Yes	Yes	No	No
Spain Madrid	CO	NO_2	O_3		PM10	SO_2	4	0 - >150	Colours	No	No	No	No
France, ATMO		NO_2	O_3	PM2.5	PM10	SO_2	3	1 - 10	Giraffe and Colours	No	No	No	No
Singapore, PSI	CO	NO_2	O_3	PM2.5	PM10	SO_2	5	0 - 500	Colours	Yes	Yes	No	No
Cairncross et al. 2007, API	CO	NO_2	O_3	PM2.5	PM10	SO_2	4	1 - 10	Colours	Yes	No	Yes	No
Stieb et al. 2008, AQHI		NO_2	O_3	PM2.5	PM10		4	1 - 10+	Colours	Yes	Yes	Yes	No
Kyrkilis et al. 2007, Aggregate AQI	CO	NO_2	O_3		PM10	SO_2						Yes	No
Sicard et al. 2011, ARI		NO_2	O_3	PM2.5	PM10	SO_2	4	0 - 10	Colours	Yes	Yes	Yes	No
Murena 2004, PI	CO	NO_2	O_3		PM10	SO_2	5	0 - 100	Clouds	No	No	Yes	Yes

Table 1: Summary of reviewed indices from government bodies and the research community

From Table 1, most indices share pollutant compositions with Ozone and Nitrogen Dioxide pollutants being common to all. They vary in the formulation of categories ranging from 3 to 6 with the mode number of categories being 4. Apart from two indices one of which colour symbology is undefined and the other one using only clouds, all other indices use colour symbology spread over their respective categories. Most indices give general and specific health recommendations to the public. The indices from the research community considered more the effect of multi-pollutants compared with the researched or implemented indices by government bodies. Most indices do not consider the location variability of pollutant measurements.

2.2 Spatial interpolation methods in air pollution

2.2.1 Spatial Interpolation

Spatial interpolation involves estimation of values of a desired attribute of a phenomenon at unsampled points using the values of sampled points in the estimation or prediction process. Miller associated Tobler's first law of geography as being core to spatial auto correlation, spatial interpolation and techniques for predicting missing variables in a geographic space (Miller, 2004). Tobler's first law of geography states that: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). In situations that involve continuous variables like elevation of the landscape, pollutant concentrations in the atmosphere and temperature, it is challenging to take measurements at every location to represent such phenomena, thus the power of spatial interpolation that facilitates the prediction of intermediate values from sampled locations and makes it possible to represent such phenomena as surfaces. Spatial interpolation is very useful in many fields like environmental modelling, surveying, mining, civil engineering, agriculture, etc., that involve the need of representing several phenomena as surfaces.

Spatial interpolation is classified into three categories of non-geostatistical, geostatistical and methods that combine both geostatistical and non-geostatistical techniques (Li and Heap, 2008). Features that distinguish and offer comparison of these spatial interpolation methods are discussed in (Li and Heap, 2014). These features include: The global and local nature of prediction where the global predictors use the entire sample points in the prediction process while the local use part of the points near the un known point for its prediction, the exactness of the interpolator where the exact interpolators resulting in a prediction that is the same as the observed value at the sampled location while the inexact results in a value that is different from the value at the sampled location, the deterministic and stochastic nature of the prediction.

2.2.2 Deterministic and stochastic spatial interpolation methods

The deterministic interpolators produce predictions without assessing the errors in the prediction process. The deterministic methods include Nearest neighbour (NN), Triangulated irregular network (TIN), Natural neighbour (NaN), Inverse distance weighting (IDW) and Radio Basis Functions (RBF) (Li and Heap, 2008; Adhikary and Dash, 2017). Stochastic interpolators produce predictions in both the deterministic part and provide the error assessment part. These include Regression models (LM) and Kriging (Li and Heap, 2008). Several researchers have used both deterministic and stochastic methods to represent different phenomena in terms of surfaces. Some studies have also compared the performance of different methods in representing different phenomena (Anselin and Le Gallo, 2006; Rojas-Avellaneda, 2007; Pultar *et al.*, 2010; Kumar Jha *et al.*, 2011; Singh *et al.*, 2011; Joseph *et al.*, 2013; Adhikary and Dash, 2017).

2.2.3 Spatial interpolation studies with air pollution

In building an environmental quality index for Madrid city, Spain, (Montero, Chasco and Larraz, 2010) used Ordinary Kriging to produce surfaces of SO_2, CO, NO_x, NO_2, PM, O_3 and noise from monitoring stations in Madrid city and linked this data with census tracts data. In their study, they reported variation of precision for locations that were further away from the monitoring stations. They used 27 monitoring stations for the annual averages of daily readings for the year 2001 for their study.

In modelling NO_2 and PM10 in the metropolitan areas of Barcelona and Bilbao, Spain, (Lertxundi-manterola and Saez, 2009) used Ordinary kriging to interpolate the daily averages of NO_2 and PM10 across the cities. They obtained a disappearing spatial dependence of concentrations at a distance of 1-3km from the monitoring stations. For the case of Barcelona, they used 9 monitoring stations for both pollutants and for the case of Bilbao, they used 28 monitoring stations for NO_2 and 10 for PM10.

In analysing the sensitivity of hedonic models of price houses with interpolation of air quality measures in Southern California, USA, (Anselin and Le Gallo, 2006) needed to assign O_3 levels to house transaction locations. They tested four interpolation methods of Thiessen polygons, IDW, Ordinary kriging and spline and found Ordinary kriging offering consistent fit and reasonable parameters for their study. They used O_3 measurements from 27 monitoring stations.

In analysing environmental justice of particulate air pollution in Hamilto, Canada, (Jerrett *et al.*, 2001) used Universal kriging to create pollution surfaces from pollutant concentration data of 23 monitoring stations and linked the results with the social economic and demographic data of Hamilton. They based on prior knowledge of existence of a spatial trend in the particulate data of Hamilton to choose universal kriging against ordinary kriging.

In predicting the pollutant concentrations in Mexico City, Mexico using interpolation methods, (Rojas-Avellaneda, 2007) identified IDW and kriging as the most used spatial interpolation methods in pollutant concentration prediction. The study compared IDW and Simple kriging methods using O_3 concentration measurements from 20 monitoring stations at a specific time for 21 days. Both methods performed better with a consideration of a linear drift in the data and produced closely related results.

In mapping the background air pollution across Europe, (Beelen *et al.*, 2009) compared the performance of three modelling techniques of Universal kriging, Ordinary kriging and a regression model. They considered pollutants of NO_2, O_3, PM10, SO_2 and CO using the Airbase data and the predictor variables used were from EU-wide databases. The results for NO_2, O_3 and PM10 showed better performance of Universal kriging compared with the other two methods while none of the methods predicted SO_2 and CO satisfactorily.

In assessing spatial interpolation methods for O_3 exposure predictions, (Joseph *et al.*, 2013) used data from two urban areas of Los Angeles, California USA, with 27 monitoring stations and Houston, Texas, with 42 monitoring stations for the compare Simple average, Nearest neighbour, IDW, Ordinary kriging and Universal kriging. The results indicated the superiority of Ordinary kriging with a calibrated range parameter in comparison with the other tested methods.

In mapping hourly O_3 episodes for spring and summer periods in Eastern Texas, USA, (Kethireddy *et al.*, 2014) used Ordinary kriging to interpolate and map O_3 at a 1km spatial scale with 80 monitoring stations. The study detected episodes of O_3 during afternoon hours and found Ordinary kriging to successful predictions of this pollutant in the study area.

2.2.4 Summary of the used spatial interpolation methods in air pollution

Table 2 shows the summary of the reviewed studies in relation to spatial interpolation and air pollution. These have been summarised with the pollutants considered and the number of used monitoring stations.

Study	Methods	Best method from comparison studies	Pollutants considered	Number of sensors
(Montero, Chasco and Larraz, 2010)	Ordinary Kriging		SO_2, CO, NO_x, NO_2, PM, O_3, noise	27
(Lertxundi-manterola and Saez, 2009)	Ordinary Kriging		NO_2, PM10	9, 28, 10
(Anselin and Le Gallo, 2006)	Thiessen polygons, IDW, Ordinary kriging and spline	Ordinary kriging	O_3	27
(Jerrett *et al.*, 2001)	Universal Kriging			23
(Rojas-Avellaneda, 2007)	IDW, Simple Kriging	IDW	O_3	20
(Beelen *et al.*, 2009)	Universal kriging, Ordinary kriging and a regression model	Universal kriging	NO_2, O_3, PM10, SO_2, CO	
(Joseph *et al.*, 2013)	Simple average, Nearest neighbour, IDW, Ordinary kriging and Universal kriging	Ordinary kriging	O_3	27, 42
(Kethireddy *et al.*, 2014)	Ordinary kriging		O_3	80

Table 2: Summary of the used spatial interpolation methods in air pollution

From these studies, we found that kriging in its different forms was the widely used method of spatial interpolation in relation to air pollution and these studies also shared the similarity in the sources of used data like the data used in Madrid with one of the studies (Montero, Chasco and Larraz, 2010) conducted in Madrid. The popularity of kriging confirms earlier findings by (Jerrett *et al.*, 2005).

2.2.5 The general prediction equation

With most prediction methods seen as weighted data averages and using the general prediction equation (Webster and Oliver, 2007), we have used this general equation (1) to briefly describe the Inverse Distance Weighted (IDW) and Kriging methods.

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{1}$$

Where $\hat{Z}$ is the predicted attribute value at x_0. $Z(x_i)$, the data values at x_i ... x_n points. λ_i are the assigned weights to the data values at those points and n is the used number of sample points for the prediction.

2.2.6 Inverse Distance Weighted (IDW)

Also known as the Inverse Distance Weighting, IDW, is a deterministic spatial interpolation method that bases on the idea of nearer things being more related to each other than further things to assign more weights to observations near the value being predicted. IDW weights the sampled points with an inverse function of the distance from the prediction point to the sampled points. Thus from the general equation (1), IDW defines weights as given in equation (2).

$$\lambda_i = \frac{1/d_i^p}{\sum_{i=1}^{n} 1/d_i^p} \tag{2}$$

Where d_i represents the distance difference between x_0 and x_i , p being the power parameter and n, the number of sample points used in the prediction. Webster and Oliver noted that the power parameter choice is arbitrary (Webster and Oliver, 2007). As cited by Li and Heap, the accuracy of IDW is mainly affected by the power parameter. With $p = 0$, IDW becomes moving average, with $p = 1$, IDW becomes linear interpolation, with $p = 2$, IDW becomes inverse square distance or inverse distance squared and when $p \neq 1$, then IDW is a weighted moving average (Li and Heap, 2008).

2.2.7 Kriging

Developed by Matheron and D.G Krige, kriging is a local, exact and stochastic spatial interpolation method. Kriging is a generic term referring to a family of geostatistical techniques that come in form of both linear or non-linear interpolators (Webster and Oliver, 2007). Kriging methods are based on the following equation (3) which is a modification from the general equation (1).

$$\hat{Z}(x_0) - \mu = \sum_{i=1}^{n} \lambda_i [Z(x_i) - \mu(x_0)] \tag{3}$$

μ, the known stationary mean is considered constant over the entire domain that is calculated as the average of the data. λ_i, kriging weight, n, number of sampled points for estimation within a search window and $\mu(x_0)$ is the mean of the sampled data within a search window.

For the purpose of this study, the focus is on the kriging methods of Ordinary kriging and Universal kriging that have been applied in similar studies more than Simple kriging.

2.2.8 Ordinary kriging (OK)

Within the kriging family, Ordinary kriging is the most used method (Webster and Oliver, 2007; Oliver and Webster, 2015). It assumes a randomly and spatially dependent variation whose mean is unknown and constant and the variance only dependent on the distance separation and direction between places (Oliver and Webster, 2015; Adhikary and Dash, 2017). The estimation error is expressed in equation (4)

$$\varepsilon(x_0) = \hat{Z}(x_0) - Z(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) - Z(x_0) \tag{4}$$

With $Z(x_0)$ as the true value of the variable at x_0 and $\varepsilon(x_0)$ as the estimation error. Since the expected residual of unbiased estimate is supposed to be 0, then

$$E[\varepsilon(x_0)] = 0 \tag{5}$$

The weights sum to 1 to ensure an unbiased estimation resulting in equation (6).

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{6}$$

2.2.9 Kriging with a trend (KT)

Also known as Universal kriging, KT considers both the random and the non-stationary nature of a variable and uses both components in the estimation process. It uses the non-stationarity component to estimate the trend and uses the random component to estimate the variogram. Thus arriving at the equation (7).

$$\hat{Z}(x_0) = \mu(x_0) + \varepsilon(x_0) \tag{7}$$

Where $\mu(x_0)$ is the deterministic function which is a drift and $\varepsilon(x_0)$ is the random variation that is treated to be auto-correlated. x_0 represents the spatial coordinates of the data as explanatory variables (Adhikary and Dash, 2017).

2.2.10 The variogram or semivariogram

A variogram expresses the spatial variation of an attribute and is estimated using a semi-variance of the difference between data values of the entire sampled points that are separated by a lag vector h. The semivariogram $\hat{\gamma}(h)$ is expressed as in equation (8).

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} [Z(x_i) - Z(x_i + h)]^2 \tag{8}$$

Where $m(h)$ is the number of sample point pairs separated by h. A plot of $\hat{\gamma}(h)$ against h gives an experimental variogram with special features of nugget, sill and range.

As cited by Li and Heap, a variogram is very important in analysing the data structure (Li and Heap, 2008). The experimental variogram produced with the sampled data is fitted with variogram models like the linear, pure nugget, spherical, exponential and Gaussian models (Bayraktar and Turalioglu, 2005; Webster and Oliver, 2007, chap. 5.2; Li and Heap, 2008). The shape of the experimental variogram determines which model is applicable for a given scenario. These fitted models help in determining the variogram parameters of nugget, sill, range, the lag size, the number of lags and the neighbourhood search strategy.

Webster and Oliver noted the controversy surrounding model choice and fitting in geostatistics due variations in different determinants like semi variances, anisotropy, point to point variation in the experimental variogram and none linearity of most models in different parameters (Webster and Oliver, 2007, p. 101).

3 Methodology

To achieve the aim and objectives of the study, we have studied the available data from Madrid city council, compared it with the reviewed air quality index approaches and interpolation methods used in the field of air pollution to find an applicable index and a suitable interpolation method for the study. We have then used these approaches during the implementation phase.

3.1 Data description, comparison with reviewed AQIs and MLAQI formulation

3.1.1 The Madrid air pollution sensor network

The city council of Madrid, has a network of 24 sensor stations deployed to measure different pollutant concentrations in Madrid city to enable pollution and air quality monitoring and management. The city council of Madrid has been monitoring air quality since 1968 using a manual network and later on a set of automatic network stations since 1978. Due to the studies, developments and legislation about air quality, the city council has continued to refine and develop this network to accommodate the developments.

The sensor stations' network continually measures the pollutants of Sulfur dioxide, Carbon monoxide, Nitrogen monoxide, Nitrogen dioxide, PM2.5, PM10, Nitrogen oxides, Ozone, Toluene, Benzene, Ethylbenzene, Metaxylene, Paraxylene, Orthoxylene, Total hydrocarbons, Methane, Non-methane hydrocarbons.

The stations in this sensor network are categorised into “Tráfico” - traffic, “Urbana de fondo” - Urban background and “Suburbana” - suburban. At each station several pollutants are measured but the combinations of measurements at each station are quite different. The tráfico sensor stations are mainly located along the road network and close to the city centre for detecting pollution caused by emissions on the road network while the other two types are located mainly outside the area covered by the tráfico sensor stations. The urbana de fondo sensors mainly represent the exposure of the general urban population while the suburbana are located in the city outskirts at locations of high ozone levels.

Stations are identified by station codes and pollutants identified by parameter codes for the pollutants measured at each station.

Tráfico - traffic stations. These are stations 4 (Pza. de España), 8 (Escuelas Aguirre), 11 (Avda. Ramón y Cajal), 36 (Moratalaz), 38 (Cuatro Caminos), 39 (Barrio del Pilar), 48 (Castellana), 50 (Plaza Castilla) and 56 (Pza. Fernández Ladreda).

Urbana de fondo - Urban background. These are stations 16 (Arturo Soria), 17 (Villaverde), 18 (Farolillo), 27 (Barajas Pueblo), 35 (Pza. del Carmen), 40 (Vallecas), 47 (Mendez Alvaro), 49 (Parque del Retiro), 54 (Ensanche de Vallecas), 55 (Urb. Embajada), 57 (Sanchinarro) and 60 (Tres Olivos Plaza)

Suburbana – suburban. These are stations 24 (Casa de Campo), 58 (El Pardo) and 59 (Juan Carlos I)

Within this network there are three “super stations” which stations measure most of the network pollutant components and consider all the types of tráfico, urbano de fondo and suburbana. These are stations 18 (Farolillo (without PM2.5)), 24 (Casa de Campo) and 8 (Escuelas Aguirre).

The location of different sensor types is shown in the Figure 1.

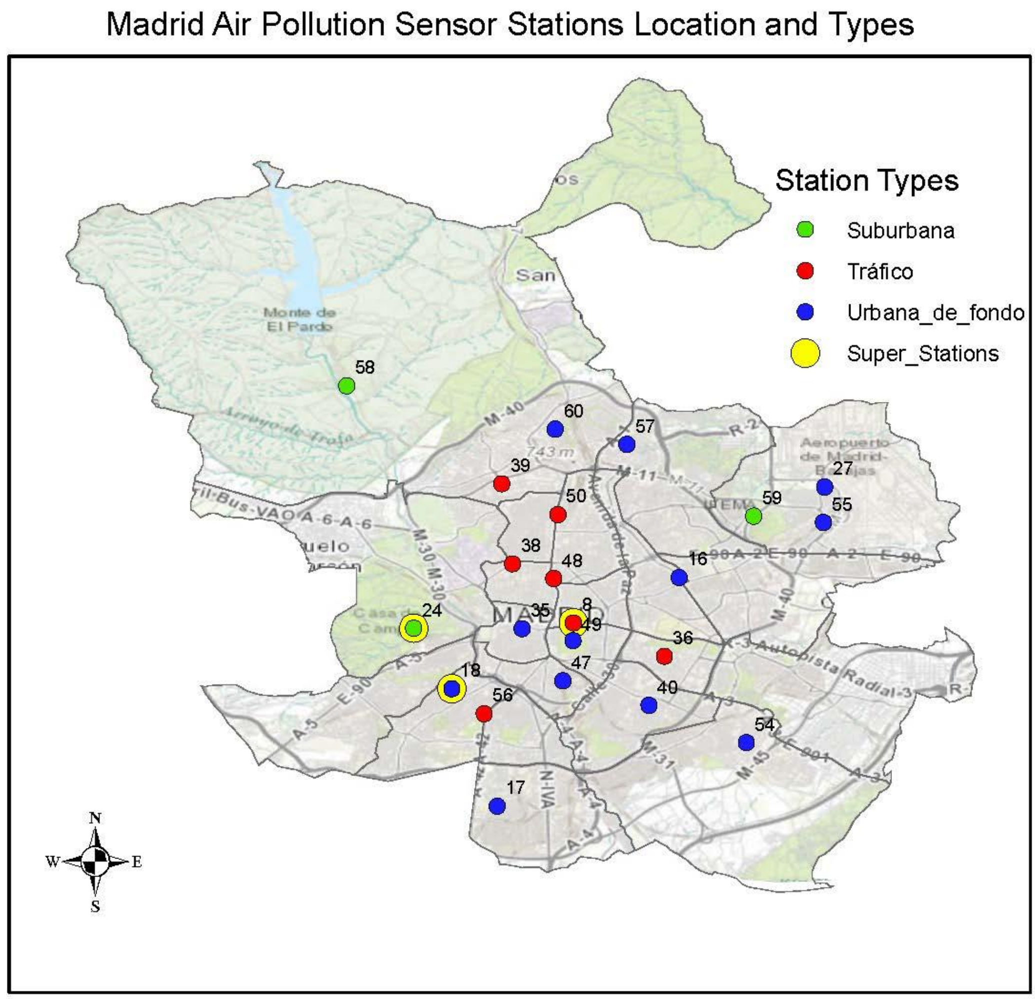

Figure 1: Sensor stations location and types

3.1.2 Description of the published data from the sensor network

The Madrid city council publishes data from the monitoring sensor stations at an hourly basis and also provides historical hourly data for different months in different file types like .XML, .CSV and .TXT file formats. The data published are the measures of pollutant concentrations. For the hourly data, the file contains about 151 records of sensors for 24 hours with the station coding, sensor coding and the date at which the values are recorded. For historical monthly data, the file contains data for the entire month where every sensor at all sensor stations has its daily hourly values.

At a single monitoring station, several sensor values for different pollutants are published but without location information. Since location information for the different sensor stations is important for spatial interpolation, which is a big component of the study, there was need to incorporate location information to the monitoring sensor stations. The Madrid city council also provides description and location (latitude, longitude and altitude) information about the network monitoring stations which we transformed into a projected coordinate system (ETRS89_UTM_zone_30N) to support structural modelling and parameter control.

3.1.3 Comparison of the data with reviewed air quality indices

In selecting an applicable air quality index for our study, we have related the available Madrid sensor data with some of the reviewed AQIs for the study. We base this relation on the indices definitions with the pollutant combinations in their formulations. From the discussed AQIs, the indices of AQHI (Canada) preferred for its linkage to health, CAQI (European), DAQI (UK Defra), Madrid Spain and ATMO (France) preferred for their formulation with the limit values of the European union have been compared with the available data from the sensor stations and this comparison is given in Table 3.

Index	**Pollutant Combinations**						**Sensor Stations**
AQHI(Canada)		Ozone	Nitrogen Dioxide			PM2.5	2
CAQI(Roadside)	carbon monoxide		Nitrogen Dioxide		PM10	PM2.5	2
CAQI(Background)	carbon monoxide	Ozone	Nitrogen Dioxide	Sulfur Dioxide	PM10	PM2.5	2
DAQI(UK Defra)		Ozone	Nitrogen Dioxide	Sulfur Dioxide	PM10	PM2.5	2
Spain Madrid	carbon monoxide	Ozone	Nitrogen Dioxide	Sulfur Dioxide	PM10		3
ATMO(France)		Ozone	Nitrogen Dioxide	Sulfur Dioxide	PM10	PM2.5	2

Table 3: Comparison of air quality indices and the available data

From Table 3, it can be seen that using the AQIs with the pollutant combinations considered during their formulations presents a challenge in interpolation as most of the AQIs are presented with 2 sensor stations that fulfil such pollutant combinations. For the AQI used by the Madrid city council, it's the three super stations that accommodate this pollutant combination though these stations are close to each other and would hardly represent the air quality situation of the whole of Madrid city.

With the main sources of pollution in Madrid being Nitrogen Dioxide mainly due to heavy traffic, ozone and PM (Madrid Salud, 2016), a couple of pollutant combinations were suggested to support interpolation of the data from the available sensor stations. Several scenarios of pollutant combination have been related and analysed with some of the reviewed AQIs in getting the optimum scenario to serve the purpose for the study.

Scenario 1. Considering the AQHI established by (Stieb *et al.*, 2008).

The AQHI defined by (Stieb *et al.*, 2008) considers either PM10 or PM2.5. Opting to use the AQHI defined by PM10 facilitates 4 candidate sensor stations from which we can interpolate the data.

However, consideration of the AQHI that uses PM2.5 facilitates 2 candidate sensor stations from which data can be interpolated. The other shortcoming of this index for this study is that it was formulated with the concentration response coefficients derived from Canadian mortality data and would not represent the situation in this study area.

Scenario 2. Considering the CAQI's Roadside index without carbon dioxide as one of the auxiliary pollutants.

Using a combination of Nitrogen Dioxide and PM10 as core pollutants with PM2.5 as an auxiliary pollutant facilitates 6 sensor stations from which data can be interpolated. Most of these stations are near the city centre of the type "Tráfico" - traffic stations. This scenario facilitates prediction of the central part of the city and not the city as a whole.

Scenario 3. Considering the CAQI's City Background index without PM10 a core pollutant and without PM2.5 as an auxiliary pollutant.

This scenario facilitates 4 sensor stations to support interpolation though these stations are close to each other and may not give a better interpolation and representation of the entire city.

Scenario 4. Considering the CAQI's City Background index without PM10 as a core pollutant and without auxiliary pollutants.

In this scenario where we only consider a combination of Nitrogen Dioxide and Ozone, it facilitates 14 sensor stations from which to interpolate the data. The challenge with this scenario is that we would neglect both PM2.5 and PM10, which are some of the pollutants of concern in the study area.

We present the analysed scenarios in Figure 2, with maps showing the available stations for interpolation in each scenario.

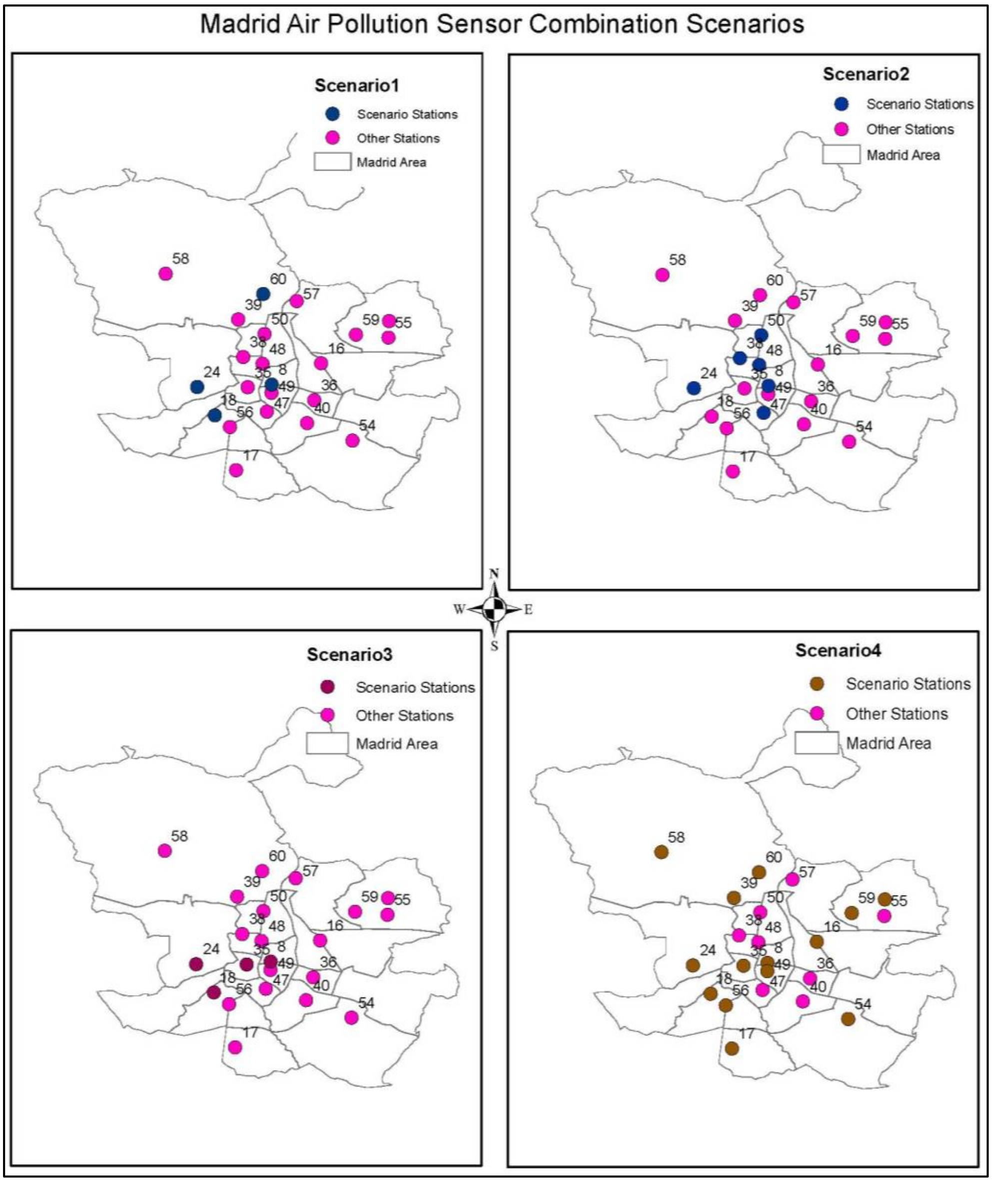

Figure 2: Madrid sensor stations combination scenarios

Though the Madrid AQI shares some features with the UK DAQI, Ireland's AQHI in terms of the pollutants and the number of categories considered, there is a difference in terms of reporting as the other two indices report daily situation rather than an hourly situation reported by the Madrid AQI. This renders it challenging to compare the limit values of these indices.

The CAQI offers both hourly and daily indices but differs from the Madrid AQI in terms of the number of categories and in their formulation. The CAQI is defined by five categories against four categories of Madrid AQI and has two types of AQI, the Roadside and Background AQIs. From the formulation of these indices and considering the hourly limit values, the limit values of NO_2 and PM10 for first two categories of CAQI for very low and low are the same as that of the first category of Madrid AQI with a difference in O_3 limits.

The Madrid AQI lacks PM2.5 in its formulation yet this pollutant is among the pollutants of concern in the city and thus the need for its inclusion in an AQI formulation.

3.1.4 Formulation of the Madrid Local Air Quality Index (MLAQI)

From this study a new hourly AQI was suggested, the Madrid Local Air Quality Index (MLAQI) which was modified out of the used index in Madrid city (*Madrid City Council*, 2015) and uses the CAQI's idea of core and auxiliary pollutants (Van Den Elshout, Léger and Heich, 2014). The MLAQI is based on the categories and limit values of the Madrid AQI and CAQI. The pollutants considered in this index are NO_2, O_3, PM10 and PM2.5. The core pollutant for MLAQI is NO_2 and the auxiliary pollutants are O_3, PM10 and PM2.5. With MLAQI, the index at a given station, should only be calculated with the existence of the core pollutant and at least one of O_3 and PM10 pollutants. This is due to the inadequacy of PM2.5 measurements and the spatial distribution of the sensors for its measurements that would not represent the whole city while interpolated.

To get a sub index, compare a pollutant concentration with the defined limit values of that pollutant and the index range for this AQI as shown in equation (9).

$$S_x = \frac{(P_x - P_{lo})}{(P_{up} - P_{lo})} * (I_{up} - I_{lo}) + I_{lo} \tag{9}$$

Where S_x is the sub index, P_x is the pollutant concentration measurement, P_{lo} is the lower limit value for the range where the pollutant measurement falls, P_{up} is the upper limit value for the range where the pollutant measurement falls, $I_{up}$is the upper limit value of the index range and I_{lo} is the lower index limit value for the range.

To get the final index at a given sensor station which qualifies for index calculation with MLAQI, use equation (9) to calculate sub indices of the available pollutants at that station. The final index is the highest of those sub indices at that station. It is the index range which defines the category of the final index. The MLAQI is shown in Table 4 with the pollutant limit values used to calculate the sub indices and the colour coding for the respective categories.

Index Range	**Index Category**	**Core Pollutant**	**Auxiliary Pollutants**		
		NO_2	**O_3**	**PM10**	**PM2.5**
0-50	Good	0 - 100	0 - 90	0-50	0-30
51-100	Acceptable	101 - 200	91 - 180	51-90	31-55
101-150	Poor	201 - 300	181 – 240	91-150	56-90
>150	Very Poor	> 300	> 240	>150	>90

Table 4: Madrid Local Air Quality Index

The strengths of this index are that it considers the local situation of Madrid city, and it considers the composition of the available sensor network and the pollutants whose concentrations can be measured. With MLAQI, the sensor network facilitates 22 out of 24 sensor stations from which we can acquire data for interpolation to represent the air quality situation in Madrid. The Figure 3 shows a map extract of the spatial distribution of the data with the defined and adopted index for the study.

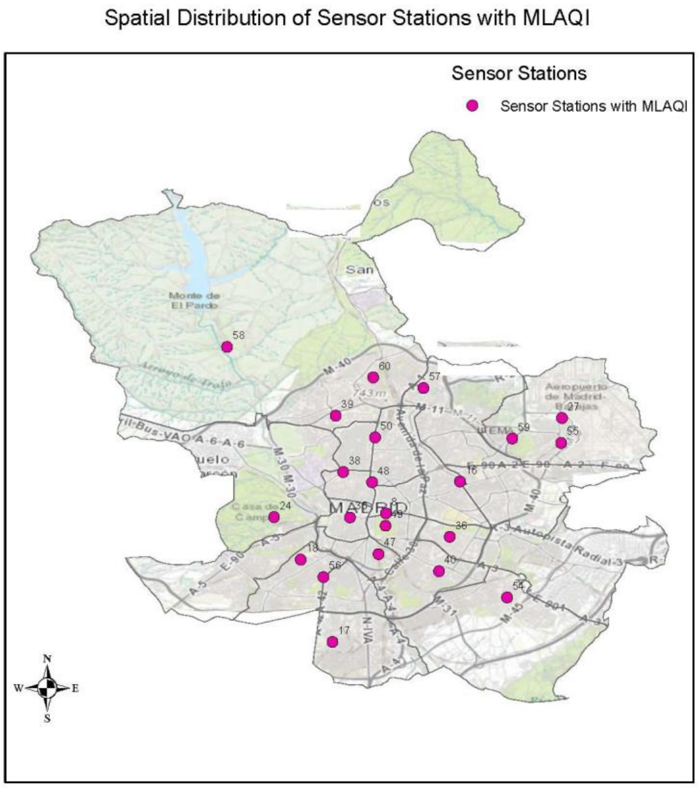

Figure 3: The spatial distribution of sensor stations with MLAQI

3.1.5 Source of the routing service

With this study using the ESRI platform for implementation and the limitations with the ESRI's routing service that limits the number of intersected streets with polygon barriers in routing, there was need for an alternative route service. We obtained this route service from the Institute of New Imaging Technologies (INIT), Universitat Jaume I (UJI) at https://geotec.init.uji.es/arcgis/rest/services/routing/SpainNetwork/NAServer/

3.2 Implementation

3.2.1 Implementation workflow

In using the described data to implement the study, we designed a workflow to guide the whole implementation procedure. This involved data acquisition from Madrid's open data portal at https://datos.madrid.es/portal/site/egob/, incorporation of the MLAQI, calculation of sub indices and the final indices, location integration and feature class generation, Exploratory Spatial Data Analysis (ESDA), structural modelling, spatial interpolation, raster-vector conversion, map processing, publishing a Web Feature Service (WFS) and creating routing web application. These steps support each other to reach our final goal. The followed workflow is graphically represented in Figure 4.

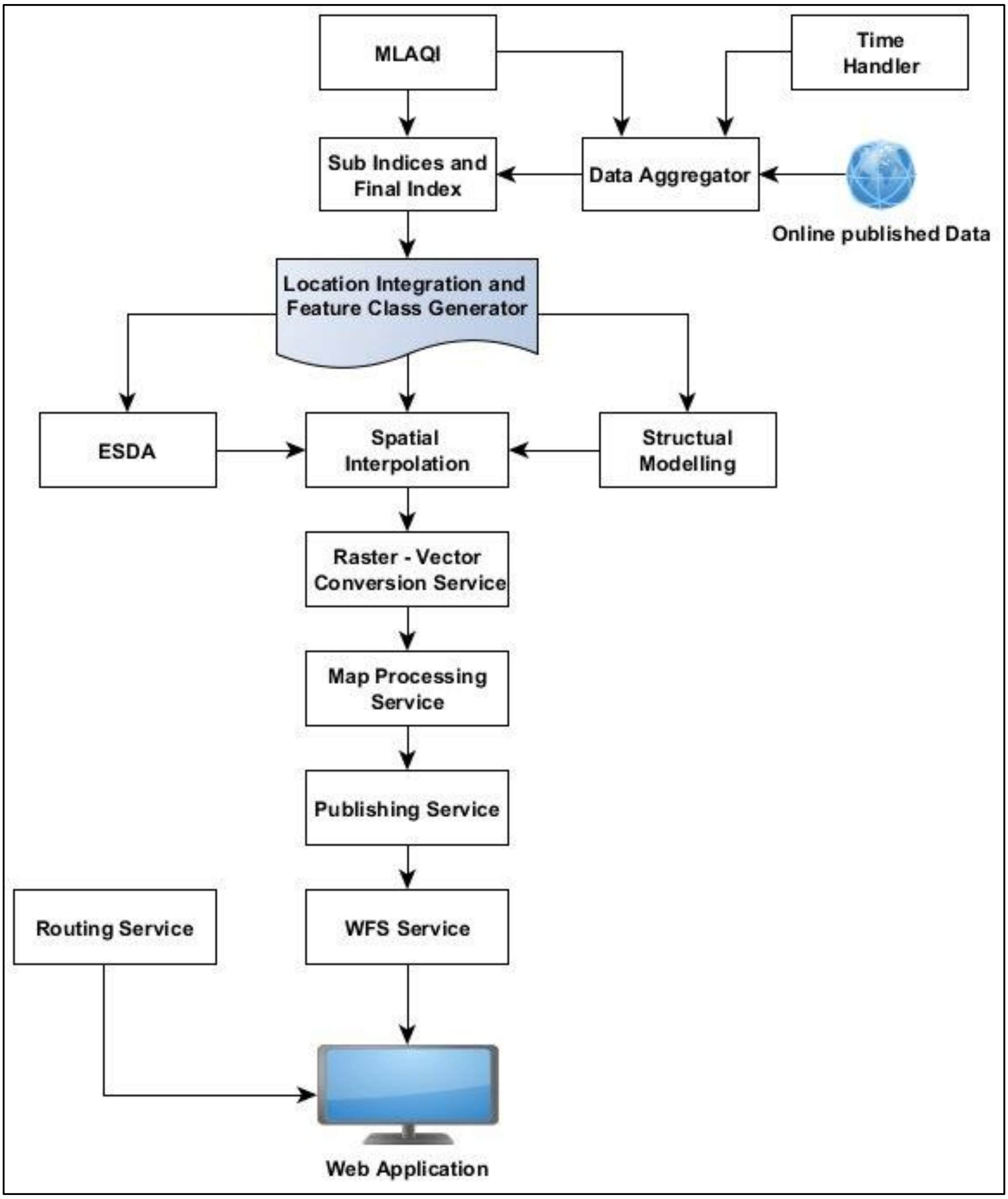

Figure 4: Implementation workflow

3.2.2 Data handling and implementation environment

For data retrieval, index calculation, ESDA and structural modelling, we identified a big challenge with manual data retrieval, handling and index calculation and a need for automation. This led us to creating several python modules for handling this challenge and automating the process. With the implementation in Python 2.7.12 under the ESRI ArcGIS ArcMap environment, having a limitation of requiring an active ArcMap sign in using the ArcGIS online account for publishing a WFS service. We opted for the implementation environment in Python 3.6.2 under ArcGIS Pro which integrates with the GIS python package for handling data and administration of ArcGIS online and ArcGIS for server.

3.2.3 Data acquisition, aggregation, index calculation and feature class generation

For each monitoring station, we grouped the sensors, and retrieved only the required pollutant data at every station. The data aggregation module retrieves real time and historical data, checks the data and extracts only the pollutants and their concentration values which are of interest in the study. The real time data uses the time handler module that returns the positions in the file at which the data should be queried. The time handler and data aggregation python modules are attached in appendices 7.1 and 7.2 respectively.

From the description of MLAQI, the index we have applied in the study, several considerations are needed for the index calculation. The consideration of core pollutant without which an index is not calculated for a given station and the need for at least one of the auxiliary pollutants. The index calculation module incorporates these conditions required for the index calculation at every monitoring station, checks pollutant concentrations against their limits, calculates the sub indices for only stations that meet the conditions and determines the final index from the highest sub index. The python code for the module is attached in appendix 7.3.

The feature class generator module joins the monitoring stations description and location information with the calculated index and then creates a point feature class for the support of processes like ESDA, structural data analysis and interpolation. The script for this process is contained in appendix 7.4.

3.2.4 Exploratory Spatial Data Analysis

Before the choice of an interpolation method for this study and interpolation, we have explored our data to check for any errors, distribution and the existence of any outliers. Since the output of the interpolation was intended to serve a continuous interpolation process all the year around for the real time data published by the Madrid city council, we decided to test 2017 historical data for diurnal and seasonal consistency in its behaviour. In the choice of time ranges for diurnal consistency analysis, we selected several hours of 7:00am, 1pm and 7pm for a specific day of 12th or 13th for the several months. For seasonal consistency analysis, we chose January, April, July and October, which are the middle months of every season. We explored the data using the regional histogram and Voronoi polygons functionalities of ArcGIS.

3.2.5 IDW structural modelling

To analyse the data structure for modelling with IDW, we used 6 sets of different model parameters to analyse our 12 datasets. For each dataset, we first used the default parameters for modelling and recorded the model parameters together with their prediction errors. To minimise the errors in relation to equations (6) and (5), we then repeated the procedure with the other 5 sets of model parameters.

3.2.6 Variogram structural modelling

To study the structure of our data for Kriging, we employed the use of the variogram and tried fitting two different models of spherical and exponential to get an optimum one to better represent our phenomenon. The exponential model appeared to fit better our phenomenon than the spherical model and was therefore used for further structural modelling of our datasets.

With the exponential model, we tried two options for all the datasets for an optimum representation. One option constituted using same model parameters for all the datasets while the other one had different model parameters for the different datasets.

With the first option using same model parameters for all the datasets except the sill, we used a nugget of 0, lag size of 2000 which was close to the Observed Mean Distance of 2417, 6 as the number of lags, 6000 as the range, 8 sectors for the sector type, 2 for minimum neighbours in a sector and 5 as the maximum neighbours in a sector. From the analysis, we recorded the values of Mean prediction error, Root Mean Square prediction error and Root Mean Square Standardised prediction error.

With the second option, we repeated the variograpy process with varying model parameters to obtain results with mean prediction error tending to 0 and Root Mean Square Standardised prediction error tending to 1. For this option, we recorded the major range, lag size, number of lags, number of neighbours, sector type and the parameters recorded in the first option.

3.2.7 Spatial interpolation of the data

With the greatly varying partial sill values identified during variogram modelling for Ordinary Kriging, which would not support the generalisation of this parameter for a continuous interpolation process for our study, we identified IDW method for application in our Madrid scenario. Using the model parameters obtained from structural modelling with IDW, we interpolated all the 12 datasets to offer the outputs of this operation as inputs for the raster vector conversion process.

3.2.8 Raster vector conversion

The input to this process is the generated geostatistical layer from the IDW interpolation. The process uses the GaLayerToContour tool using the filled contour type with the class breaks of the MLAQI index to generate an index categorised polygon feature class. The extent environment settings for this process are set using the extent of the Madrid boundary polygon feature class. This is used to extend the processing environment outside the location of the sensor locations' point data to generate representation for the entire area of Madrid. Using the clip analysis tool, the generated polygon feature class is now clipped with the Madrid boundary polygon feature class to keep it within the shape extent of the study area. The output from the clipping is projected to a WGS1984 Web Mercator Auxiliary Sphere coordinate system to support the drawing of the features on the web. Using the select analysis tool, the projected polygon feature class is separated into 4 polygon feature classes according to the MLAQI index.

3.2.9 Map processing service

This process is important to prepare the map elements for publishing. Feature classes in a geodatabase do not store symbology with them, thus the need for this step to apply the polygon outlines, colour and transparency symbology properties before publishing. This process mainly uses the mapping module of ArcPy. The process checks the layer's list in the project's map document for existence of any layers and removes them. It then adds the created index feature classes to the map. It then uses the apply symbology from layer tool to apply the symbology to the layers in the map and saves the map document. The map processing module is attached in appendix 7.5.

3.2.10 Publishing service

The publishing service involves steps of creating a web layer service definition draft (SDDraft), staging the service definition, connecting to the server, adding or updating the service difinition, publishing the service on the server and setting sharing options on the server. The service name, summary and tags variables are used to describe the service once it is published. The Create Web Layer SDDraft mapping function takes in the parameters of the prepared map document, the SDDraft path, the service name, the hosting destination, tags and summary of the service to create the SDDraft file. The SDDraft is staged with the StageService_server function to create the Service Definition file (SD). Using the GIS python package, the module then connects to the server with the publisher or administrator credentials to add a new service definition or update the existing one on the server. The module then publishes the service definition on the server setting overwriting and sharing options. The publishing module is attached in appendix 7.6.

3.2.11 Deployment of an hourly services execution

To achieve the hourly publishing service, we used Microsoft Window's Task Scheduler application. The steps required for the Task Scheduler process were creating a process name, creating a trigger and creating an action to be performed at that trigger. The trigger includes the time, the process should run and after what interval the process should repeat. We created an execution task to run the python execution file with an argument of the deployment service python file. This python file contains the required python modules for processing all the steps until the publishing is executed.

3.2.12 Web application development

With the created WFS service and routing service as data sources, we created a web application using the ArcGIS API for JavaScript technology. The application uses the JavaScript references for CSS and JavaScript from ESRI's Content Delivery Network (CDN). The other component for the application is the HTML5. With the ArcGIS API for JavaScript built on top of the Dojo framework, we used some of its controls and layout elements for the application. The involved steps to create the application were creating the HTML5 document with basic HTML tags of html, head, meta, tittle, body. We then added CSS styles for the used layout elements. The main part of the application is the script part that queries and adds the data to the application and controls the process and the behaviour of the application.

With the ArcGIS JavaScript API and Dojo using Asynchronous Module Definition (AMD), this allows only calling appropriate modules for the application with Dojo's require function. This "require" function takes two arguments: one for the required modules array and the other one, a function for returning those loaded modules. We created a map containing a basemap and the route layer, the map view containing the map, data extent and the spatial reference of the data. We then added all the WFS service layers using the feature layer module.

The application queries the poor and very poor feature layers from the WFS service in a JSON format using the Query module, checks if those layers contain features in them and adds such features to the route parameters as polygon barriers. This is to check for empty layers which when added to the route parameters block routing execution in the application. These polygon barriers are used to limit the user navigation in these high pollution areas. A route generated with a polygon barrier between the locations of interest will minimise the travelling cost parameter and avoid these areas with high pollution index categories.

To add stops on the map, the user clicks the Add Stop button which activates the adding stop functionality. The user then adds stops on the map view by clicking on the desired locations on the view.

The route task uses the route layer and the route parameters of stops, polygon barriers, impedance attribute name and spatial reference to determine the route. To obtain and display a route between at least points of interest selected on the map, the user clicks the solve route button which determines and displays the shortest route through those points while minimising polluted area. The used travel impedance for the study is kilometres thus the resulting shortest route.

We enhance the user's interaction with the application by providing the user with more functionalities with the application. These include the clear stops and route button to enable the user clear the route result and perform another route task, the search bar offering the user with functionality to search for a location of interest, a geolocation button to enable the user locate their position on the map, the zoom and home buttons for map navigation, the legend for identifying index categories and the basemap toggle tool to enable the user toggle the base the streets and hybrid basemaps.

4.0 Results and discussion

4.1 Statistics from ESDA

The following are the results from ESDA for all the datasets of the study.

January data

For the three selected hours on 12th January, data from 22 out of 24 sensor stations met the requirements for the MLAQI calculation and their indices calculated. Using the regional histogram, obtained statistics for the calculated indices.

For the 7am dataset, we obtained a minimum index value of 4 at Juan Carlos I station and a maximum index value of 73 at Plaza Fernández Ladreda station. The index distribution was slightly skewed to the right with a mean of 32 slightly greater than a median of 30. There was a deviation of 15 from the mean value. The 1st quartile indicated that 25% of the stations had indices less than an index value of 23 and the 3rd quartile indicated that 75% of the stations had indices less than an index value of 39. With Plaza Fernández Ladreda station having the maximum index value of 73 that did not pass the 1.5IQR (1.5 inter quartile rule) test for the upper outlier, this station was further investigated and this value was due to the high NO_2 value of 145 recorded at that time and there were other readings close to 130 after 4 hours.

For the 1pm dataset, we obtained a minimum index value of 12 at Juan Carlos I station and a maximum index value of 61 at Cuatro Caminos station. The index distribution was slightly skewed to the left with the mean of 37 slightly less than the median of 39. There was a deviation of 13 from the mean value. The 1st quartile indicated that 25% of the stations had indices less than 28 while the 3rd quartile indicated that 75% of the stations had indices less than 47. There was no identified outlier in this dataset.

For the 7pm dataset, the obtained minimum index value of 14 was at Juan Carlos I station and the maximum index value of 37 was at Escuelas Aguirre station. We obtained a normal index distribution with the same mean and median of 28. There was a deviation of 5 from the mean value. About 25% of the stations had indices less than 26 and 75% of the stations had indices less than 31. With Juan Carlos I station having the minimum value of 14 that appeared as a lower outlier, the station was further investigated and found the contributing pollutants of NO_2 and O_3 for index calculation had stable readings compared with their neighbouring readings on that day.

Generally, the analysis with Voronoi polygons did not indicate any presence of trends in the datasets for January. The index values for 7am and 7pm datasets spread with in the first two categories of the MLAQI while the 7pm dataset only stayed in the first category of the MLAQI.

April

For the case of April, data was queried at the same time intervals on the 13th day of the month. Exactly 22 out of 24 stations passed the MLAQI criteria for index calculation in all the three cases and their indices calculated.

The 7am dataset produced a minimum index value of 15 at El Pardo station and a maximum index value of 59 at Plaza Fernández Ladreda station. This dataset showed a normal index distribution with the same mean and median of 33. There was a deviation of 9 from the mean value. The 1st

quartile indicated that 25% of the station had indices less than 28 while the third quartile indicated that 75% of the stations had indices less than 38. The maximum index value of 59 was further investigated and the contributing pollutant NO_2 had a value of 117 that was more than 40 comparing to its neighbouring values.

The 1pm dataset showed a minimum index value of 10 at Sanchinarro station and a maximum index value of 70 at El Pardo station. The index distribution was skewed to the left with a mean of 48 and a median of 58. There was a deviation of 19 from the mean value. Around 25% of the stations had index values less than 27 while 75% of the stations had index values less than 63.

The 7pm dataset showed a minimum index value of 11 at Sanchinarro station and a maximum index value of 74 at Casa de Campo station. The index distribution was skewed to the left with a mean of 49 and a median of 58. There was a deviation of 19 from the mean value. 25% of the stations had index values less than 35 while 75% of the stations had index values less than 63.

Generally, there was no indication of trends in the datasets for April. All the three data sets are spread in the first two categories of the MLAQI.

July data

For the July datasets, the data was obtained for the 13th day of the month. The 7am dataset had 22 out of 24 stations that qualified for the index calculation whereas the 1pm and 7pm datasets had 21 out of 24 stations that qualified for the index calculation.

The 7am dataset had a minimum index value of 15 at Parque del Retiro station and a maximum index value of 71 at Cuatro Caminos station. The index had a normal distribution with the same mean and median of 37. There was a deviation of 12 from the mean value. The statistics showed that 25% of the stations had index values less than 29 while 75% of the stations had index values less than 44. We further investigated the station with the maximum index value and found that the contributing pollutant PM10 had a reading of 67 that was stable with its neighbouring readings.

With the 1pm dataset, the minimum index value was 30 at Méndez Álvaro station and the maximum index value was 158 at Urbanización Embajada station. The index had a slight skewed distribution to the right with a mean of 57 and a median value of 52. The deviation from the mean was 26. The statistics showed that 25% of the stations had index values less than 45 while 75% of the stations had index values less than 61. We tested the maximum index value with the 1.5IQR rule and it indicated to be an upper outlier. The Urbanización Embajada station with this maximum index value was further investigated and the contributing pollutant of PM10 had a reading of 168 that was more than 80 compared with its neighbouring values. We also noted that this index value fell in the very poor category of MLAQI.

With the 7pm dataset, the minimum index value was 31 at Barrio del Pilar station and the maximum index value was 102 at Sanchinarro station. The index distribution was skewed to the right with a mean value of 57 and a median value of 47. There was a deviation of 22 from the mean value. 25% of the stations had index values less than 41 while 75% of the stations had index values less than 65. We further investigated Sanchinarro station and the contributing pollutant PM10 had a reading of 92, which was stable compared with its neighbouring readings.

With the July datasets, there appeared index values that were higher than the rest of the values. These were investigated further to have a better understanding of their distribution. Apart from the maximum index value of the 1pm dataset, the other index values in this dataset are spread out within the first two categories of the MLAQI like the 7am data set while the 7pm dataset slightly extends to the third category of the MLAQI. There were no noticeable tendencies of trend in the datasets.

October data

With the datasets obtained for October 12th, 22 out of 24 stations for all the three datasets qualified for the MLAQI criteria for index calculation and their index values calculated.

With the 7am dataset, the minimum index value was 22 at Ensanche de Vallecas station while the maximum value was 47 at Plaza Fernández Ladreda station. The index distribution was normal with a mean of 34 and a median of 33. There was a deviation of 7 from the mean value. The statistics show that 25% of the stations had index values less than 27 while 75% of the stations had index values less than 38. There were no noticeable outliers in this dataset.

For the 1pm dataset, the minimum was 26 at Méndez Álvaro station while the maximum was 52 at Casa de Campo station. The index distribution was normal with mean value of 40 and median value of 39. There was a deviation of 7 from the mean value. The statistics revealed that 25% of the stations had index values less than 37 while 75% of the stations had index values less than 46. There were no noticeable outliers in the dataset.

For the 7pm dataset, the minimum value was 29 at Plaza Castilla station and the maximum value 93 at Escuelas Aguirre station. The index distribution is skewed to the right with a mean value of 56 and a median value of 49. There was a deviation of 19 from the mean value. Around 25% of the stations had index values less than 40 while 75% of the stations had index values less than 66. There were no noticeable outliers in the dataset.

Generally, there were no outlier indications in the October datasets. The 7am dataset is spread only in the first category of the MLAQI while the 1pm and 7pm datasets spread out in the first and second categories of MLAQI. There were no trends noticed in the datasets.

4.1.1 General nature of the data

Generally, most of the data from all the months spread in the first two categories of the MLAQI. The stations whose index values appeared to be outliers were further investigated and the values of the contributing pollutants to index calculation compared with their neighbouring values. We identified the index values that appeared to be outliers for January and April were as a result of NO_2 while those that appeared to be outliers for July were as a result of PM10. It should however be noted that Madrid city normally gets high pollution episode days when pollutant concentrations reach above 200 or 300 especially with NO_2. In 2017, there were NO_2 episodes in March, September, October, November and December. We captured part of the October episode on 12th October with NO_2 reading for Escuelas Aguirre station as 185 at 7pm. The next reading at 8pm was 259 and the reading at 9pm was 228. No trend was detected from the datasets. The statistics from the datasets are summarised in Table 5.

Month(2017)	Time	Stations	Minimum	Maximum	Mean	Standard Deviation	1st Quartile	Median	3rd Quartile	MLAQI Categories	Lower Outlier	Upper Outlier
January 12th												
	7am	22	4	73	32	15	23	30	39	2	-1	63
	1pm	22	12	61	37	13	28	39	47	2	-0.5	75.5
	7pm	22	14	37	28	5	26	28	31	1	18.5	38.5
April 13th												
	7am	22	15	59	33	9	28	33	38	2	13	53
	1pm	22	10	70	48	19	27	58	63	2	-27	117
	7pm	22	11	74	49	19	35	58	63	2	-7	105
July 13th												
	7am	22	15	71	37	12	29	37	44	2	6.5	66.5
	1pm	21	30	158	57	26	45	52	61	4	21	85
	7pm	21	31	102	57	22	41	47	65	3	5	101
Oct 12th												
	7am	22	22	47	34	7	27	33	38	1	10.5	54.5
	1pm	22	26	52	40	7	37	39	46	2	23.5	59.5
	7pm	22	29	93	56	19	40	49	66	2	1	105

Table 5: Summary statistics of Exploratory Spatial Data Analysis

4.2 Statistics from IDW structural modelling

Table 6 shows a pair extract of model parameters together with their errors that produced the lowest magnitude Mean and Root Mean Square prediction errors. From Table 6, the model parameters of power 2, search radius of 6000 for both the major and minor semi-axes, minimum neighbours 5, maximum neighbours 10 and sector type of 8 sectors, produced the highest frequency of 67% in obtaining the lowest magnitude Root Mean Square prediction error. With this consistency in producing lowest prediction errors, we selected these model parameters for further interpolation with IDW.

Month (2017)	Time	Power	Search Radius	Minimum Neighbours	Maximum Neighbours	Sector Type	Minimum	Maximum	Average	Mean Prediction Error	Root Mean Square Prediction Error
January 12th											
	7am	2	6000	5	10	8 Sectors	4	73	32	1.0564	16.9887
		2	3000	3	10	8 Sectors	4	73	32	0.9605	17.1555
	1pm	2	6000	5	10	8 Sectors	12	61	37	2.1592	15.0591
		2	2000	2	6	8 Sectors	12	61	37	2.0599	15.4057
	7pm	2	3000	3	10	8 Sectors	14	37	28	0.7371	6.0215
		2	2000	2	6	8 Sectors	14	37	28	0.788	6.0126
April 13th											
	7am	2	3000	3	10	8 Sectors	15	59	33	1.3371	9.5417
		2	2000	2	6	8 Sectors	15	59	33	1.2907	9.5561
	1pm	2	6000	5	10	8 Sectors	10	70	48	-2.1524	19.8638
		2	4500	2	10	4 Sectors	10	70	48	-1.6527	20.5964
	7pm	2	6000	5	10	8 Sectors	11	74	49	-2.3625	18.5848
		2	4500	2	10	4 Sectors	11	74	49	-1.8538	19.7176
July 13th											
	7am	2	6000	5	10	8 Sectors	15	71	37	0.873	13.2113
		2	3000	3	10	8 Sectors	15	71	37	0.7485	13.3889
	1pm	2	6000	5	10	8 Sectors	30	158	57	-1.4705	28.283
		2	4500	2	10	4 Sectors	30	158	57	-0.6452	29.2997
	7pm	2	6000	5	10	8 Sectors	31	102	57	-0.5447	24.0025
		2	2000	2	6	8 Sectors	31	102	57	0.081	24.9419
Oct 12th											
	7am	2	6000	5	10	4 Sectors at 45^{o}	22	47	34	-0.1571	5.7084
		2	4500	2	10	4 Sectors	22	47	34	-0.0629	5.9538
	1pm	2	6000	5	10	8 Sectors	26	52	40	-1.3401	7.1394
		2	2000	2	6	8 Sectors	26	52	40	-1.2768	7.1233
	7pm	2	6000	5	10	8 Sectors	29	93	56	1.3273	22.7593
		2	3000	3	10	8 Sectors	29	93	56	0.9549	23.1866

Table 6: An extract of model parameters and their errors with IDW structural modelling

4.3 Statistics from variogram structural modelling

The results from structural modelling with same model parameters except the sill are presented in Table 7.

Month(2017)	Time	Partial sill	Minimum	Maximum	Average	Mean Prediction Error	Root Mean Square Prediction Error	Root Mean Square Standardized Prediction Error
January 12th								
	7am	214.72	4	73	32	0.0114	16.6543	1.2744
	1pm	192.76	12	61	37	1.2719	15.2679	1.2142
	7pm	31.566	14	37	28	0.33	5.8922	1.2378
April 13th								
	7am	79.58	15	59	33	0.4296	9.1827	1.1379
	1pm	415.98	10	70	48	-1.2642	19.0342	0.9888
	7pm	392.55	11	74	49	-1.2733	18.1727	0.9648
July 13th								
	7am	168.43	15	71	37	0.3989	13.3432	1.2291
	1pm	693.41	30	158	57	-1.1333	26.7615	1.125
	7pm	510.53	31	102	57	0.1917	24.536	1.1824
Oct 12th								
	7am	53.86	22	47	34	-0.2058	6.4456	0.905
	1pm	53.81	26	52	40	-0.6958	6.791	0.9838
	7pm	421.47	29	93	56	0.4155	23.8407	1.4116

Table 7: Error summary from variogram using same model parameters

From the results in Table 7, the magnitude of the mean prediction error for the datasets which had very high or very low indices in them is higher and their Root Mean Square prediction error values are higher compared to the other datasets. From equations (6) and (5), the Mean prediction error is expected to be close to 0 and Root Mean Squared Standardised prediction error close to 1 but with these results not tending very close to those values, we carried out the second option and compared the results of both options.

We present the obtained results from the second option using varying model in Table 8.

Month(2017)	Time	Major Range	Partial Sill	Lag size	Number of lags	Minimum Neighbours	Maximum Neighbours	Sector Type	Minimum	Maximum	Average	Mean Prediction Error	Root Mean Square Prediction Error	Root Mean Square Standardized Prediction Error
January 12th														
	7am	5850	244.98	2650	8	2	5	8 sectors	4	73	32	0.0099	16.65099	1.1843
	1pm	2500	185.39	500	6	2	5	8 sectors	12	61	37	0.7039	15.0139	1.064
	7pm	2500	28.77	600	6	2	5	8 sectors	14	37	28	0.0446	5.7011	1.0515
April 13th														
	7am	4000	81.87	1000	6	2	5	8 sectors	15	59	33	0.0632	9.3366	1.0453
	1pm	2700	352.23	900	6	2	5	8 sectors	10	70	48	-0.9333	20.1774	1.0354
	7pm	4000	315.48	800	6	2	5	8 sectors	11	74	49	-1.08464	18.7351	1.04342
July 13th														
	7am	3900	192.93	800	8	2	5	8 sectors	15	71	37	0.0872	13.0541	1.0022
	1pm	5500	655.36	1650	8	2	5	8 sectors	30	158	57	-1.1338	26.6378	1.1282
	7pm	4000	470.71	1650	8	2	5	8 sectors	31	102	57	0.0104	24.4779	1.1326
Oct 12th														
	7am	9900	51.28	1650	6	2	5	4 sectors	22	47	34	0.0044	6.0834	1.0089
	1pm	3900	45.44	650	6	2	5	8 sectors	26	52	40	-0.4683	6.9245	1.0173
	7pm	4000	479.47	1000	6	2	5	8 sectors	29	93	56	-0.0054	23.1574	1.145

Table 8: Error summary from variogram using varying model parameters

From Table 8, the values for the mean prediction error and the Root Mean Squared Standardised prediction error tended closer to 0 and 1, respectively, compared with the first option results. Root Mean Square prediction error did not greatly differ from that of the first option. Like in the first option, the datasets with very high or very low index values exhibited the same behaviour.

Though the results from both options were closely related, there was a challenge of greatly varying partial sill values for the first option with the same model parameters. This made it challenging to support an hourly continuous representation of our index for different diurnal times and seasons.

4.4 Categorised surfaces from IDW interpolation

We categorised the output from IDW interpolation of all datasets according to the definition of MLAQI. We present the output from this process in Figure 5.

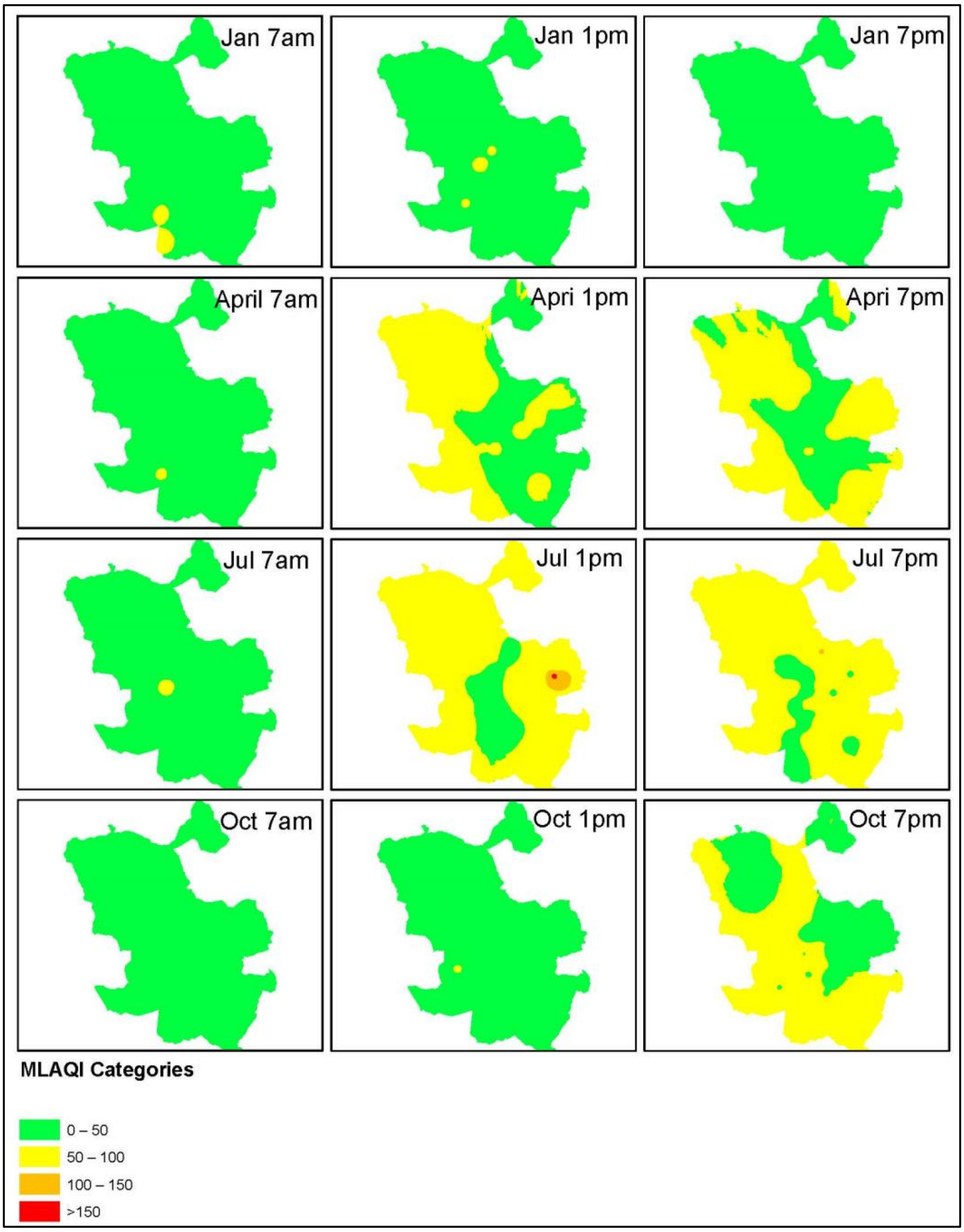

Figure 5: Categorised IDW interpolation output according to MLAQI

From Figure 5, the categorised surfaces mostly use the first two categories of MLAQI. The datasets for 7am data for all the months are mostly spread in the good category of MLAQI while those of 7pm tend to spread in the good and acceptable categories. The data from January is mostly depicted by the good category of MLAQI. The data from the months of April and July are more represented by the acceptable category for the afternoon hours. These results agree with the findings obtained during ESDA.

4.5 The published Web Feature Service (WFS)

The output from the publishing operation is a WFS service. This service resides under the ArcGIS REST Services Directory in the folder specified during publishing. The main features of the created service are the creation of a Service ItemId (b813a7936d194bc1bfc078b6c91ba29a) that can be used for its querying, maintaining the layer organisation, contains the Spatial Reference ID (SRID) for the data, gives the supporting query formats of both JSON and geoJSON and also maintains the symbology properties which are contained in the drawing info parameter. This service is shared publicly and can be reached using the ItemId or the REST service URL at [https://services1.arcgis.com/k8WRSCmxGgCwZufI/arcgis/rest/services/madridPollutionSurface/FeatureServer.](https://services1.arcgis.com/k8WRSCmxGgCwZufI/arcgis/rest/services/madridPollutionSurface/FeatureServer) The service definition for this service is overwritten every hour and the service then updated every hour from this service definition.

4.6 Resulting routing application

The developed application is hosted at http://pixel.uji.es/pollution/. This application uses the hourly generated features from the feature service URL, which it continuously refreshes to fetch any updates and apply them to the map. A screenshot of the application is shown in Figure 6.

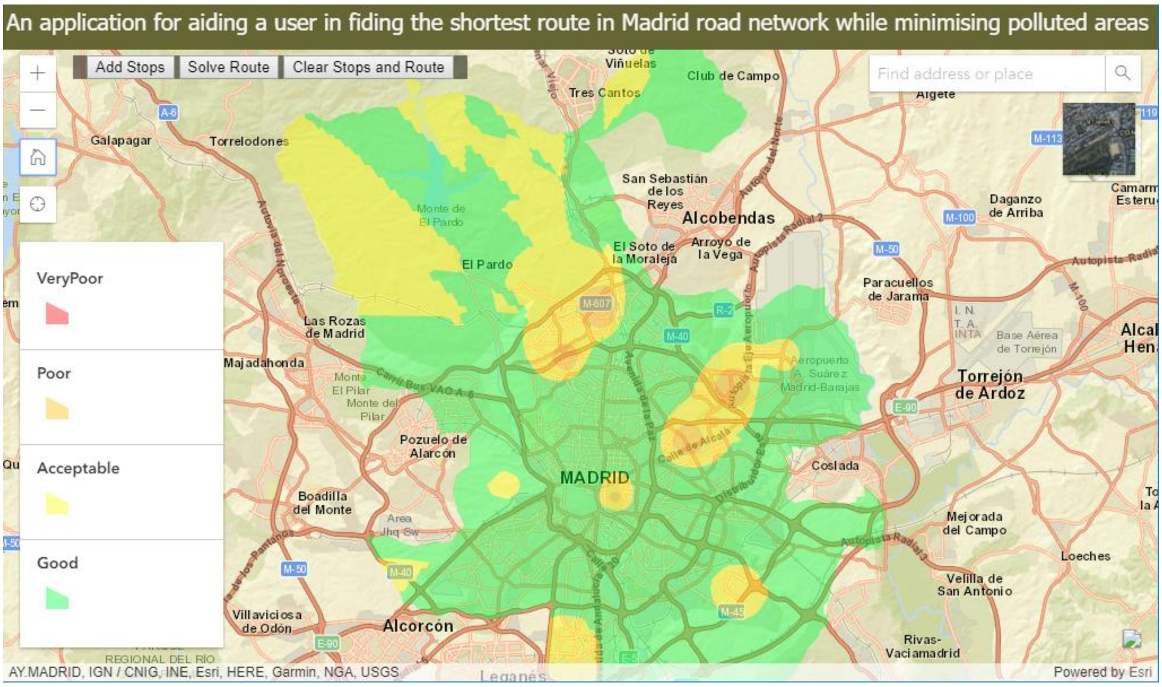

Figure 6: A screenshot of the developed application

To protect the user's health against polluted area of poor and very poor categories, the application offers the user with functionality to solve a route for navigation through Madrid road network which minimises these polluted areas. A screenshot of such a rout is shown in Figure 7.

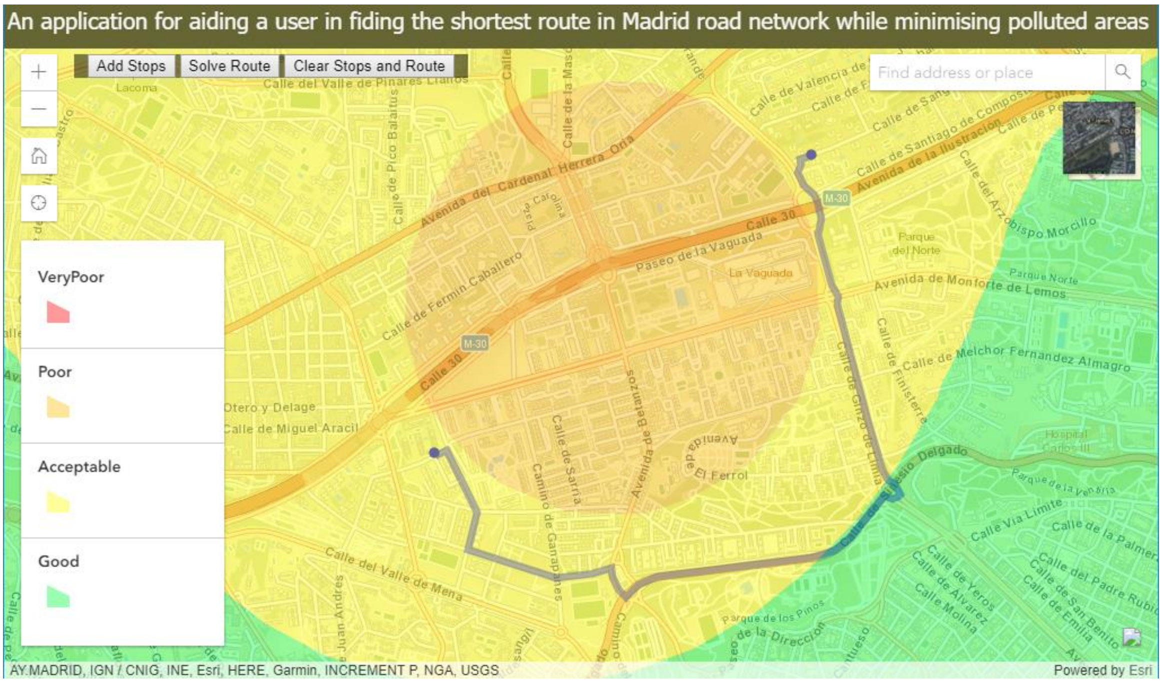

Figure 7: A route with minimised polluted areas

5.0 Conclusion and recommendations

The Madrid pollution sensor station network infrastructure provides a platform to help in building services aimed at helping the public be aware of the air quality around them.

Using this network's data and modifying the city's air quality index for better representation of the pollution phenomenon in the city, the study has developed a hourly pollution surface service to help the public be aware of the air quality around them and be in position to make informed decisions while planning their hourly activities.

The study also demonstrated that the created service could be used as a data source by other applications to create applications aimed at public awareness of the air pollution around them.

During the implementation of the study, each process appeared to be in position to execute standalone operations for other purposes of mapping and data analysis like automating feature class generation, mapping and publishing automations.

From this study, we made an observation that data will not always be in a ready geospatial format for individual studies, but geospatial technologies are enablers to extract and format such data to serve the purpose of such studies.

With the challenge of varying sill values during the variogram modelling, further research could be performed with more spatio-temporal analysis of hourly behaviour of the pollution situation in Madrid. This would help us have a better understanding about episode hours in Madrid and whether hourly varying interpolation model parameters could be applicable to achieve better interpolation.

More studies could incorporate mortality or hospital admissions data in the air quality index and results compared.

The interpolation could be supplemented with other parameters like elevation of sensor stations to test the variability of pollutant concentration measurements over different elevation.

A routing application could be extended into a mobile application with detailed study of how people would interact with such an application.

6.0 References

Adhikary, P. P. and Dash, C. J. (2017) 'Comparison of deterministic and stochastic methods to predict spatial variation of groundwater depth', *Applied Water Science*, 7(1), pp. 339–348. doi: 10.1007/s13201-014-0249-8.

Anselin, L. and Le Gallo, J. (2006) 'Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial Aspects', *Spatial Economic Analysis*. Taylor & Francis Group, 1(1), pp. 31–52. doi: 10.1080/17421770600661337.

ATMO France (2008) *Indice de Qualité d'Air*. Available at: http://www.atmo-france.org/fr/index.php?/2008043044/indice-de-qualite-d-air/id-menu-275.html (Accessed: 23 October 2017).

Bayraktar, H. and Turalioglu, F. S. (2005) 'A Kriging-based approach for locating a sampling site - In the assessment of air quality', *Stochastic Environmental Research and Risk Assessment*, 19(4), pp. 301–305. doi: 10.1007/s00477-005-0234-8.

Beelen, R. *et al.* (2009) 'Mapping of background air pollution at a fine spatial scale across the European Union', *Science of the Total Environment*, 407(6), pp. 1852–1867. doi: 10.1016/j.scitotenv.2008.11.048.

Cairncross, E. K., John, J. and Zunckel, M. (2007) 'A novel air pollution index based on the relative risk of daily mortality associated with short-term exposure to common air pollutants', *Atmospheric Environment*, 41(38), pp. 8442–8454. doi: 10.1016/j.atmosenv.2007.07.003.

COMEAP UK (2011) *Review of the UK Air Quality Index*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/304633/COMEAP_review_of_the_uk_air_quality_index.pdf (Accessed: 18 October 2017).

Correia, A. W. *et al.* (2013) 'Effect of Air Pollution Control on Life Expectancy in the United States', *Epidemiology*, 24(1), pp. 23–31. doi: 10.1097/EDE.0b013e3182770237.

ECCC Canada (2016) *About the Air Quality Health Index*. Available at: http://www.ec.gc.ca/cas-aqhi/default.asp?lang=En&n=065BE995-1 (Accessed: 20 October 2017).

Van Den Elshout, S., Léger, K. and Heich, H. (2014) 'CAQI Common Air Quality Index — Update with PM 2.5 and sensitivity analysis', *Science of the Total Environment*, 488–489, pp. 461–468. doi: 10.1016/j.scitotenv.2013.10.060.

van den Elshout, S., Léger, K. and Nussio, F. (2008) 'Comparing urban air quality in Europe in real time. A review of existing air quality indices and the proposal of a common alternative', *Environment International*, 34(5), pp. 720–726. doi: 10.1016/j.envint.2007.12.011.

Emily Connolly *et al.* (2013) *Update on Implementation of the Daily Air Quality Index*. Available at: https://uk-air.defra.gov.uk/assets/documents/reports/cat14/1304251155_Update_on_Implementation_of_the_DAQI_April_2013_Final.pdf (Accessed: 20 October 2017).

EPA Ireland (2013a) *Development of the Air Quality Index for Health (AQIH)*. Available at: http://www.epa.ie/pubs/reports/air/quality/DevelopmentOfAQIH2013v3.0.pdf (Accessed: 20 October 2017).

EPA Ireland (2013b) *Air Quality Index for Health*. Environmental Protection Agency (EPA). Available at: http://www.epa.ie/air/quality/index/ (Accessed: 21 October 2017).

Jerrett, M. *et al.* (2001) 'A GIS - Environmental justice analysis of particulate air pollution in

Hamilton, Canada', *Environment and Planning A*, 33(6), pp. 955–973. doi: 10.1068/a33137.

Jerrett, M. *et al.* (2005) 'A review and evaluation of intraurban air pollution exposure models', *Journal of Exposure Analysis and Environmental Epidemiology*, 15(2), pp. 185–204. doi: 10.1038/sj.jea.7500388.

Joseph, J. *et al.* (2013) 'Application of validation data for assessing spatial interpolation methods for 8-h ozone or other sparsely monitored constituents', *Environmental Pollution*, 178, pp. 411–418. doi: 10.1016/j.envpol.2013.03.035.

Kethireddy, S. R. *et al.* (2014) 'Geospatial interpolation and mapping of tropospheric ozone pollution using geostatistics', *International Journal of Environmental Research and Public Health*, 11(1), pp. 983–1000. doi: 10.3390/ijerph110100983.

Kumar Jha, D. *et al.* (2011) 'Evaluation of Interpolation Technique for Air Quality Parameters in Port Blair, India', *Universal Journal of Environmental Research and Technology*, 1(3), pp. 301–310. Available at: http://www.environmentaljournal.org/1-3/ujert-1-3-10.pdf (Accessed: 13 December 2017).

Kyrkilis, G., Chaloulakou, A. and Kassomenos, P. A. (2007) 'Development of an aggregate Air Quality Index for an urban Mediterranean agglomeration: Relation to potential health effects', *Environment International*, 33(5), pp. 670–676. doi: 10.1016/j.envint.2007.01.010.

Lertxundi-manterola, A. and Saez, M. (2009) 'Modelling of nitrogen dioxide (NO 2) and fine particulate matter (PM 10) air pollution in the metropolitan areas of Barcelona and', *Environmetrics*, (2), pp. 477–493. doi: 10.1002/env.939.

Li, J. and Heap, A. D. (2008) 'A Review of Spatial Interpolation Methods for Environmental Scientists', *Geoscience Australia*, (Record 2008/23), p. 137. doi: http://www.ga.gov.au/image_cache/GA12526.pdf.

Li, J. and Heap, A. D. (2014) 'Spatial interpolation methods applied in the environmental sciences: A review', *Environmental Modelling and Software*, 53, pp. 173–189. doi: 10.1016/j.envsoft.2013.12.008.

Madrid City Council (2015) *Air Quality Information System*. Available at: http://www.mambiente.munimadrid.es/opencms/opencms/calaire/SistemaIntegral/SistInformacion.html (Accessed: 23 October 2017).

Madrid Salud (2016) *Health surveillance of environmental risks*. Available at: http://madridsalud.es/vigilancia-sanitaria-de-riesgos-ambientales/#fusion-tab-saludycontaminación (Accessed: 4 December 2017).

Miller, H. J. (2004) 'Tobler's First Law and Spatial Analysis', *Annals of the Association of American Geographers*. Blackwell Publishing, 94(2), pp. 284–289. doi: 10.1111/j.1467-8306.2004.09402005.x.

Montero, J.-M., Chasco, C. and Larraz, B. (2010) 'Building an environmental quality index for a big city: a spatial interpolation approach combined with a distance indicator', *Journal of Geographical Systems*, 12(4), pp. 435–459. doi: 10.1007/s10109-010-0108-6.

Murena, F. (2004) 'Measuring air quality over large urban areas: Development and application of an air pollution index at the urban area of Naples', *Atmospheric Environment*, 38(36), pp. 6195–6202. doi: 10.1016/j.atmosenv.2004.07.023.

NEA Singapore (2014) *Computation of the Pollutant Standards Index (PSI)*. Available at: http://www.haze.gov.sg/docs/default-source/faq/computation-of-the-pollutant-standards-index-(psi).pdf (Accessed: 23 October 2017).

NEA Singapore (2017) *Air Quality Information (PSI)*. Available at: http://www.haze.gov.sg/air-quality-information (Accessed: 23 October 2017).

Oliver, M. A. and Webster, R. (2015) *Basic Steps in Geostatistics: The Variogram and Kriging*. Cham: Springer International Publishing (SpringerBriefs in Agriculture). doi: 10.1007/978-3-319-15865-5.

Plaia, A. and Ruggieri, M. (2011) 'Air quality indices: a review', *Reviews in Environmental Science and Bio/Technology*, 10(2), pp. 165–179. doi: 10.1007/s11157-010-9227-2.

Pultar, E. *et al.* (2010) 'EDGIS: a dynamic GIS based on space time points', *International Journal of Geographical Information Science*. Taylor & Francis , 24(3), pp. 329–346. doi: 10.1080/13658810802644567.

Rojas-Avellaneda, D. (2007) 'Spatial interpolation techniques for estimating levels of pollutant concentrations in the atmosphere', *Revista Mexicana de Fisica*, 53(6), pp. 447–454. Available at: http://www.scielo.org.mx/pdf/rmf/v53n6/v53n6a4.pdf (Accessed: 18 December 2017).

Sicard, P. *et al.* (2011) 'Air quality trends and potential health effects - Development of an aggregate risk index', *Atmospheric Environment*, 45(5), pp. 1145–1153. doi: 10.1016/j.atmosenv.2010.12.052.

Singh, V. *et al.* (2011) 'A cokriging based approach to reconstruct air pollution maps, processing measurement station concentrations and deterministic model simulations', *Environmental Modelling and Software*, 26(6), pp. 778–786. doi: 10.1016/j.envsoft.2010.11.014.

Stieb, D. M. *et al.* (2008) 'A new multipollutant, no-threshold air quality health index based on short-term associations observed in daily TimeSeries analyses', *Journal of the Air & Waste Management Association*, 58(3), pp. 435–450. doi: 10.3155/1047-3289.58.3.435.

Tobler, W. R. (1970) 'A Computer Movie Simulating Urban Growth in the Detroit Region', *Economic Geography*, 46, p. 234. doi: 10.2307/143141.

US EPA (2013) *Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI)*. Available at: https://www3.epa.gov/airnow/aqi-technical-assistance-document-may2016.pdf (Accessed: 13 October 2017).

US EPA (2014) *Air Quality Index - A Guide to Air Quality and Your Health*. Available at: https://www3.epa.gov/airnow/aqi_brochure_02_14.pdf (Accessed: 29 September 2017).

Valarm (2018) *Air Quality and Pollution Monitoring with Industrial IoT Sensor for Measuring PM1, PM2.5, PM10, O3, CH4, H2S, VOCs, SO2, NOx, NO, NO2*. Available at: http://www.valarm.net/air-quality-monitoring-with-iot-pollution-sensors-like-ozone-sulfur-nitrous-oxides-dust-particulates-pm2-5/ (Accessed: 16 February 2018).

Webster, R. (Richard) and Oliver, M. A. (2007) *Geostatistics for Environmental Scientists.* Available at: https://www.soils.org/publications/vzj/abstracts/1/2/321.

Zanella, A. *et al.* (2014) 'Internet of Things for Smart Cities', *IEEE Internet of Things Journal*, 1(1), pp. 22–32. doi: 10.1109/JIOT.2014.2306328.

7.0 Appendices

7.1 The time handler module

```
# -*- coding: utf-8 -*-

# The time Handler Module

# checking to provide index positions of values at the current time(tm)

def currHour(tm):
    # conditions for 24 hours time variation
    if tm >= 0 and tm <= 24:
        if tm >= 0.34 and tm < 1.34:
            # dVI data value index, vTI index of the data validity test
            dVI, vTI = 55, 56
        elif tm >= 1.34 and tm < 2.34:
            dVI, vTI = 9, 10
        elif tm >= 2.34 and tm < 3.34:
            dVI, vTI = 11, 12
        elif tm >= 3.34 and tm < 4.34:
            dVI, vTI = 13, 14
        elif tm >= 4.34 and tm < 5.34:
            dVI, vTI = 15, 16
        elif tm >= 5.34 and tm < 6.34:
            dVI, vTI = 17, 18
        elif tm >= 6.34 and tm < 7.34:
            dVI, vTI = 19, 20
        elif tm >= 7.34 and tm < 8.34:
            dVI, vTI = 21, 22
        elif tm >= 8.34 and tm < 9.34:
            dVI, vTI = 23, 24
        elif tm >= 9.34 and tm < 10.34:
            dVI, vTI = 25, 26
        elif tm >= 10.34 and tm < 11.34:
            dVI, vTI = 27, 28
        elif tm >= 11.34 and tm < 12.34:
            dVI, vTI = 29, 30
        elif tm >= 12.34 and tm < 13.34:
            dVI, vTI = 31, 32
        elif tm >= 13.34 and tm < 14.34:
            dVI, vTI = 33, 34
        elif tm >= 14.34 and tm < 15.34:
            dVI, vTI = 35, 36
        elif tm >= 15.34 and tm < 16.34:
            dVI, vTI = 37, 38
        elif tm >= 16.34 and tm < 17.34:
            dVI, vTI = 39, 40
        elif tm >= 17.34 and tm < 18.34:
            dVI, vTI = 41, 42
        elif tm >= 18.34 and tm < 19.34:
            dVI, vTI = 43, 44
        elif tm >= 19.34 and tm < 20.34:
            dVI, vTI = 45, 46
        elif tm >= 20.34 and tm < 21.34:
            dVI, vTI = 47, 48
        elif tm >= 21.34 and tm < 22.34:
            dVI, vTI = 49, 50
        elif tm >= 22.34 and tm < 23.34:
            dVI, vTI = 51, 52
        else:
            dVI, vTI = 53, 54
        tIndixValues = [dVI, vTI]
    return tIndixValues
```

7.2 Data aggregator module

```
# Data aggregation module

# -*- coding: utf-8 -*-
import time
import timeFcDec


def aggrData(a):
    # storing all data into a single list to be returned
    storePollData = []
    # using the local time
    currTimeLocal = time.localtime()
    decTime = (currTimeLocal[3] + (float(currTimeLocal[4])/60))
    dataIndexHolder = timeFcDec.currHour(decTime)
    # extracting lines from the text file
    for lineData in a:
        if lineData[dataIndexHolder[1]] == 'V':
            processPart = (lineData[0]+lineData[1]+lineData[2]+"
"+lineData[3]+" " +
                                    lineData[4]+" "+lineData[5]+" " +
lineData[6]+" " +
                                    lineData[7]+" "+lineData[8]+"
"+lineData[dataIndexHolder[0]])
            storePollData.append(processPart)

    # testing the station and component pollutants
    stnNosRef = ['004', '008', '011', '016', '017', '018', '024', '027',
'035',
                        '036', '038', '039', '040', '047', '048', '049',
'050',
                        '054', '055', '056', '057', '058', '059', '060']
    pollNos = ['08', '09', '10', '14']
    # storing all component pollutant data into respective stations
    stnNos = ['004', '008', '011', '016', '017', '018', '024', '027',
'035', '036',
                            '038', '039', '040', '047', '048', '049',
'050', '054',
                            '055', '056', '057', '058', '059', '060']
    # the length of the entire list
    dataLength = len(storePollData)
    i = 0
    for i in range(dataLength):
        m = storePollData[i].split()
        if m[0][-3:] in stnNosRef and m[1] in pollNos:
            r = stnNosRef.index(m[0][-3:])
            if r >= 0:
                stnNos[r] += (" "+m[1]+" "+m[7])
    return stnNos
    pass
```

7.3 The sub index and final index calculation module

```
# Index calculation module

# Ilo - Index lower limit, Iup - Index upper limit
# Plo - pollutant lower limit, Pup - pollutant upper limit
# Px -  pollutant value whose index we want, Pco - pollutant code

# indexLimits = [0, 50, 51, 100, 101, 150, >150]

# function for determining nitrogen params for index
def nitLim(Px):
    # nitrogenLimitValues = [0, 100, 101, 200, 201, 300]
    # conditions for Pco 08(Nitrogen dioxide)
    if Px >= 0 and Px <= 100:
        Ilo, Iup, Plo, Pup = 0, 50, 0, 100
    elif Px > 100 and Px <= 200:
        Ilo, Iup, Plo, Pup = 51, 100, 101, 200
    elif Px > 200 and Px <= 300:
        Ilo, Iup, Plo, Pup = 101, 150, 201, 300
    else:
        # arbitrary addition of 100 to the pollutant reading to allow index calc
        # arbitrary Iup of 200 for index calc
        Ilo, Iup, Plo, Pup = 151, 200, 301, Px+100
    sx = ((Px - Plo)/(Pup - Plo))*(Iup - Ilo) + Ilo
    return sx

# function for determining ozone params for index
def ozoLim(Px):
    # ozoneLimitValues = [0, 90, 91, 180, 181, 240]
    # conditions for Pco 14(Ozone)
    if Px >= 0 and Px <= 90:
        Ilo, Iup, Plo, Pup = 0, 50, 0, 90
    elif Px > 90 and Px <= 180:
        Ilo, Iup, Plo, Pup = 51, 100, 91, 180
    elif Px > 180 and Px <= 240:
        Ilo, Iup, Plo, Pup = 101, 150, 181, 240
    else:
        # arbitrary addition of 100 to the pollutant reading to allow index calc
        # arbitrary Iup of 200 for index calc
        Ilo, Iup, Plo, Pup = 151, 200, 241, Px+100
    sx = ((Px - Plo)/(Pup - Plo))*(Iup - Ilo) + Ilo
    return sx

# function for determining pm10 params for index
def ptnLim(Px):
    # pmtenLimitValues = [0, 50, 51, 90, 91, 150]
    # conditions for Pco 10(PM10)
    if Px >= 0 and Px <= 50:
        Ilo, Iup, Plo, Pup = 0, 50, 0, 50
    elif Px > 50 and Px <= 90:
        Ilo, Iup, Plo, Pup = 51, 100, 51, 90
    elif Px > 90 and Px <= 150:
        Ilo, Iup, Plo, Pup = 101, 150, 91, 150
    else:
        # arbitrary addition of 100 to the pollutant reading to allow index calc
        # arbitrary Iup of 200 for index calc
        Ilo, Iup, Plo, Pup = 151, 200, 151, Px+100
    sx = ((Px - Plo)/(Pup - Plo))*(Iup - Ilo) + Ilo
    return sx
```

```python
# function for determining pm2.5 params for index
def ptfLim(Px):
    # ptwodotfiveLimitValues = [0, 30, 31, 55, 56, 90]
    # conditions for Pco 09(PM2.5)
    if Px >= 0 and Px <= 30:
        Ilo, Iup, Plo, Pup = 0, 50, 0, 30
    elif Px > 30 and Px <= 55:
        Ilo, Iup, Plo, Pup = 51, 100, 31, 55
    elif Px > 55 and Px <= 90:
        Ilo, Iup, Plo, Pup = 101, 150, 56, 90
    else:
        # arbitrary addition of 100 to the pollutant reading to allow index calc
        # arbitrary Iup of 200 for index calc
        Ilo, Iup, Plo, Pup = 151, 200, 91, Px+100
    sx = ((Px - Plo)/(Pup - Plo))*(Iup - Ilo) + Ilo
    return sx

def dataIndices(a):
    finalDataList = []
    # a is the whole data list
    # navigate through the entire data
    wdl = len(a)  # the length of the whole data
    i = 0
    for i in range(wdl):
        stnWithIndex = []
        b = a[i].split()  # b is a station data list
        # navigate through a station data
        sdl = len(b)  # the length of the station data
        if "08" in b and sdl > 3:  # condition for core pollutant
            if "14" in b or "10" in b:  # condition for auxilliary pollutants
                stnWithIndex.append(b[0])
                subIndicesList = []
                j = 1
                for j in range(sdl):
                    # nitrogen a core pollutant, is already in the condition
                    indexN = b.index("08")
                    indexNV = indexN+1
                    nitSubIndex = nitLim(float(b[indexNV]))
                    subIndicesList.append(nitSubIndex)
                    # test existance of ozone in the station list
                    if "14" in b:
                        indexO = b.index("14")
                        indexOV = indexO+1
                        ozoSubIndex = ozoLim(float(b[indexOV]))
                        subIndicesList.append(ozoSubIndex)
                    # test existance of pm10 in the station list
                    if "10" in b:
                        indexT = b.index("10")
                        indexTV = indexT+1
                        ptnSubIndex = ptnLim(float(b[indexTV]))
                        subIndicesList.append(ptnSubIndex)
                    # test existance of pm25 in the station list
                    if "09" in b:
                        indexF = b.index("09")
                        indexFV = indexF+1
                        ptfSubIndex = ptfLim(float(b[indexFV]))
                        subIndicesList.append(ptfSubIndex)
                    pass
                stnIndex = int(round(max(subIndicesList)))
                stnWithIndex.append(stnIndex)
                finalDataList += stnWithIndex
        pass
    return finalDataList
    pass
```

7.4 Feature class generation, Interpolation and vector conversion processes

```
# Location integration, Feature class generation, Interpolation and Vector
conversion
# -*- coding: utf-8 -*-
import urllib.request
import os
import arcpy
import madpoldataaggregCT  # calling the aggregation module
import madpolindexcalculation  # calling the index calculation module


arcpy.env.workspace = "C:\\MadridPollution\\ToneDev\\Madrid_Sensors.gdb"
arcpy.env.overwriteOutput = True

# local variables for feature class creation
out_path = arcpy.env.workspace
out_name = "testmadrid"
geometry_type = "POINT"
has_m = "DISABLED"
has_z = "DISABLED"
# projected coordinate system used for the area around madrid
sr = arcpy.SpatialReference(25830)
# Create empty point Featureclass
arcpy.CreateFeatureclass_management(out_path, out_name, geometry_type, "",
has_m, has_z, sr)

empData = out_name


# adding required fields
def addRFields(featLyr):
    nFieldName = "stnID"
    if len(arcpy.ListFields(featLyr, nFieldName)) == 0:
        arcpy.AddField_management(featLyr, nFieldName, "TEXT")
    nFieldName = "stnName"
    if len(arcpy.ListFields(featLyr, nFieldName)) == 0:
        arcpy.AddField_management(featLyr, nFieldName, "TEXT")
    nFieldName = "stnType"
    if len(arcpy.ListFields(featLyr, nFieldName)) == 0:
        arcpy.AddField_management(featLyr, nFieldName, "TEXT")
    nFieldName = "stnElev"
    if len(arcpy.ListFields(featLyr, nFieldName)) == 0:
        arcpy.AddField_management(featLyr, nFieldName, "SHORT")
    nFieldName = "MLAQI"
    if len(arcpy.ListFields(featLyr, nFieldName)) == 0:
        arcpy.AddField_management(featLyr, nFieldName, "SHORT")


# check if fields are added
addRFields(empData)
```

```
stnLoc = [['004', 'Plaza de España', 439577.503, 4475070.366, 637, 'Urbana de tráfico'],
    ['008', 'Escuelas Aguirre', 442117.68, 4474786.082, 672, 'Urbana de tráfico'],
    ['011', 'Ramón y Cajal', 442567.611, 4478088.964, 708, 'Urbana de tráfico'],
    ['016', 'Arturo Soria', 445788.017, 4476804.759, 698, 'Urbana de fondo'],
    ['017', 'Villaverde', 439420.566, 4466527.998, 601, 'Urbana de fondo'],
    ['018', 'Farolillo', 437893.845, 4471845.515, 581, 'Urbana de fondo'],
    ['024', 'Casa de Campo', 436601.241, 4474586.797, 645, 'Suburbana'],
    ['027', 'Barajas Pueblo', 450835.949, 4480855.369, 631, 'Urbana de fondo'],
    ['035', 'Plaza del Carmen', 440344.765, 4474535.745, 657, 'Urbana de fondo'],
    ['036', 'Moratalaz', 445254.758, 4473251.64, 671, 'Urbana de tráfico'],
    ['038', 'Cuatro Caminos', 440036.799, 4477465.145, 699, 'Urbana de tráfico'],
    ['039', 'Barrio del Pilar', 439693.285, 4481085.951, 673, 'Urbana de tráfico'],
    ['040', 'Vallecas', 444705.01, 4471040.179, 677, 'Urbana de fondo'],
    ['047', 'Méndez Álvaro', 441734.865, 4472167.017, 609, 'Urbana de fondo'],
    ['048', 'Castellana', 441455.003, 4476789.943, 676, 'Urbana de tráfico'],
    ['049', 'Parque del Retiro', 442095.972, 4473986.767, 662, 'Urbana de fondo'],
    ['050', 'Plaza Castilla', 441622.45, 4479674.926, 728, 'Urbana de tráfico'],
    ['054', 'Ensanche de Vallecas', 448055.403, 4469329.231, 630, 'Urbana de fondo'],
    ['055', 'Urbanización Embajada', 450784.046, 4479256.762, 619, 'Urbana de fondo'],
    ['056', 'Plaza Fernández Ladreda', 439003.502, 4470704.937, 605, 'Urbana de tráfico'],
    ['057', 'Sanchinarro', 444026.471, 4482834.185, 700, 'Urbana de fondo'],
    ['058', 'El Pardo', 434382.749, 4485555.056, 616, 'Suburbana'],
    ['059', 'Juan Carlos I', 448379.789, 4479557.516, 669, 'Suburbana'],
    ['060', 'Tres Olivos', 441559.491, 4483537.375, 715, 'Urbana de fondo']]

# Adding points and updating the fields
fields = ["SHAPE@XY", "stnID", "stnName", "stnType", "stnElev", "MLAQI"]

#       Fetching Data from the madrid city council URL
source = "http://www.mambiente.munimadrid.es/opendata/horario.txt"
with urllib.request.urlopen(source) as response:
    byteData = response.read().decode('utf-8') # decode from byte format
# store the decoded data in a 2D list
rawData =[]
kmt = byteData.split()
for k in kmt:
    rawData.append(k.split(","))

pldata = madpoldataaggregCT.aggrData(rawData)
dWI = madpolindexcalculation.dataIndices(pldata)

def usedStnData(a, b):
    # retrieve only stations with MLAQI measurements
    stnLocUsed = []
    kt = len(a)
    i = 0
    for i in range(kt):
        if a[i][0] in b:
            listVal = a[i]
            stnLocUsed.append(listVal)
            i += 1
    # update the used stations with the index
    ktused = len(stnLocUsed)
    j = 0
    for j in range(ktused):
        indIndex = b.index(stnLocUsed[j][0])+1
        stnLocUsed[j].append(b[indIndex])
        j += 1
    # print stnLocUsed
    return stnLocUsed
    pass
```

```
# organise the data according to feature class structure using draftData
draftData = []
shpData = usedStnData(stnLoc, dWI)
for rowStn in shpData:
    ptLoc = (rowStn[2], rowStn[3])
    ptID = rowStn[0]
    ptNam = rowStn[1]
    ptTy = rowStn[5]
    ptEl = rowStn[4]
    ptInd = rowStn[6]
    ptdec = (ptLoc, ptID, ptNam, ptTy, ptEl, ptInd)
    draftData.append(ptdec)
# insert the data into the feature class
with arcpy.da.InsertCursor(empData, fields) as updLyr:
        for row in draftData:
            updLyr.insertRow(row)

#            IDW Interpolation
# local variables for IDW interpolation
in_feat = "testmadrid"
zField = "MLAQI"
out_feat = "surfacemadrid"
IDW_Layer = "IDW_Layer"
Output_raster = ""
cellSize = 65
power = 2
# search neighborhood
majSemiaxis = 6000
minSemiaxis = 6000
angle = 0
maxNeighbors = 10
minNeighbors = 3
sectorType = "EIGHT_SECTORS"
searchNeighbourhood = arcpy.SearchNeighborhoodStandard(majSemiaxis, min-
Semiaxis,
                                                       angle, maxNeighbors,
                                                       minNeighbors, sec-
torType)
# Enable Geostatistical Analyst license
arcpy.CheckOutExtension("GeoStats")
# IDW interpolation
arcpy.IDW_ga(in_feat, zField, IDW_Layer, Output_raster, cellSize, power,
searchNeighbourhood)
# Creating pollution shapes from interpolation
tempEnvironment0 = arcpy.env.extent
arcpy.env.extent = "424753.6621 4462565.9412 456039.9542 4499364.5676"
arcpy.GALayerToContour_ga(IDW_Layer, "FILLED_CONTOUR", out_feat, "DRAFT",
"MANUAL", "", "50;100;150;200")
arcpy.env.extent = tempEnvironment0
# Delete the IDW_Layer
arcpy.Delete_management(IDW_Layer)
```

```
# Clipping local variables
clip_infeat = out_feat
clip_feat = "DISTRITOS"
clip_outfeat = "clippedFeat"
xy_tolerance = ""
# Clipping out the required extents
arcpy.Clip_analysis(clip_infeat, clip_feat, clip_outfeat, xy_tolerance)
# reproject Features
clipProjected = "clipFeatProj"
arcpy.Project_management("clippedFeat", "clipFeatProj", "PRO-
JCS['WGS_1984_Web_Mercator_Auxiliary_Sphere',GEOGCS['GCS_WGS_1984',DA-
TUM['D_WGS_1984',SPHE-
ROID['WGS_1984',6378137.0,298.257223563]],PRIMEM['Greenwich',0.0],UNIT['De-
gree',0.0174532925199433]],PROJECTION['Mercator_Auxiliary_Sphere'],PARAME-
TER['False_Easting',0.0],PARAMETER['False_Northing',0.0],PARAMETER['Cen-
tral_Meridian',0.0],PARAMETER['Standard_Parallel_1',0.0],PARAMETER['Auxil-
iary_Sphere_Type',0.0],UNIT['Meter',1.0]]", "ETRS_1989_To_WGS_1984", "PRO-
JCS['ETRS_1989_UTM_Zone_30N',GEOGCS['GCS_ETRS_1989',DA-
TUM['D_ETRS_1989',SPHE-
ROID['GRS_1980',6378137.0,298.257222101]],PRIMEM['Greenwich',0.0],UNIT['De-
gree',0.0174532925199433]],PROJECTION['Transverse_Mercator'],PARAME-
TER['False_Easting',500000.0],PARAMETER['False_Northing',0.0],PARAME-
TER['Central_Meridian',-3.0],PARAMETER['Scale_Factor',0.9996],PARAME-
TER['Latitude_Of_Origin',0.0],UNIT['Meter',1.0]]", "NO_PRESERVE_SHAPE", "",
"NO_VERTICAL")
# Add barrier attribute fields
def addRFields(featLyr):
    nFieldName = "Name"
    if len(arcpy.ListFields(featLyr, nFieldName)) == 0:
        arcpy.AddField_management(featLyr, nFieldName, "TEXT")
    nFieldName = "BarrierType"
    if len(arcpy.ListFields(featLyr, nFieldName)) == 0:
        arcpy.AddField_management(featLyr, nFieldName, "SHORT")
    nFieldName = "Attr_TravelTime"
    if len(arcpy.ListFields(featLyr, nFieldName)) == 0:
        arcpy.AddField_management(featLyr, nFieldName, "FLOAT")
# check if fields are added
addRFields(clipProjected)

barfields = ["Classes", "Name", "BarrierType", "Attr_TravelTime"]

with arcpy.da.UpdateCursor(clipProjected, barfields) as updbLyr:
        for row in updbLyr:
            row[2] = 1
            if row[0]==0:
                row[1],row[3] = "Good Category", 0.25
            elif row[0]==1:
                row[1],row[3] = "Acceptable Category", 0.75
            elif row[0]==2:
                row[1],row[3] = "Poor Category", 1.25
            elif row[0]==3:
                row[1],row[3] = "Very Poor Category", 1.75
            # update all the row fields
            updbLyr.updateRow(row)
# seperate the shape into seperate index categories
sepFeat = clipProjected
arcpy.Select_analysis(sepFeat, "Good", "Classes = 0")
arcpy.Select_analysis(sepFeat, "Acceptable", "Classes = 1")
arcpy.Select_analysis(sepFeat, "Poor", "Classes = 2")
arcpy.Select_analysis(sepFeat, "VeryPoor", "Classes = 3")
print ("This is the last index category")
```

7.5 Map processing service

```python
# Map processing service

# -*- coding: utf-8 -*-
import arcpy
import os

arcpy.env.overwriteOutput = True

# Referencing the project and its contents
projLoc = 'C:\\MadridPollution\\ToneDev\\'
aprx = arcpy.mp.ArcGISProject(projLoc+"ToneDev.aprx")
aprxMap = aprx.listMaps("Map")[0]

# remove existing layers from the map inside the project
lyrs = aprxMap.listLayers()
for lyr in lyrs:
    aprxMap.removeLayer(lyr)

# add the updated layers to the map inside the project
aprxMap.addDataFromPath(projLoc + "Madrid_Sensors.gdb\\VeryPoor")
aprxMap.addDataFromPath(projLoc + "Madrid_Sensors.gdb\\Poor")
aprxMap.addDataFromPath(projLoc + "Madrid_Sensors.gdb\\Acceptable")
aprxMap.addDataFromPath(projLoc + "Madrid_Sensors.gdb\\Good")


# Reference symbology sources
symb1 = projLoc + "Good.lyrx"
symb2 = projLoc + "Acceptable.lyrx"
symb3 = projLoc + "Poor.lyrx"
symb4 = projLoc + "VeryPoor.lyrx"

# apply symbology to the current layers in the document
doclyrs = aprxMap.listLayers()
arcpy.ApplySymbologyFromLayer_management(doclyrs[0], symb1)
arcpy.ApplySymbologyFromLayer_management(doclyrs[1], symb2)
arcpy.ApplySymbologyFromLayer_management(doclyrs[2], symb3)
arcpy.ApplySymbologyFromLayer_management(doclyrs[3], symb4)

aprx.save()
del aprx
```

7.6 Publishing service

```
# Publishing service

# -*- coding: utf-8 -*-
import arcpy, os
from arcgis.gis import GIS

arcpy.env.workspace = os.path.join('C:\\', 'MadridPollution', 'ToneDev')
arcpy.env.overwriteOutput = True
# location of the map document
wrkspc = 'C:\\MadridPollution\\ToneDev\\'
serviceName = "madridPollutionSurface"
portal = "http://www.arcgis.com" # Can also reference a local portal
user = "twesigye_uji"
password = "y20tplt18jabs"
# Set sharing options
shrOrg = True
shrEveryone = True
shrGroups = ""

# paths for the required files
SDdraft = os.path.join(wrkspc, serviceName + ".sddraft")
SD = os.path.join(wrkspc, serviceName + ".sd")
thumbn  = os.path.join(wrkspc, "MLAQI.jpg")

# provide summary and tags of the service
summary = 'Pollution surface for madrid city generated from MLAQI'
tags = 'Public health, Madrid city, air pollution, air quality'

# delete the SDDraft and SD files if they exist on the system.
if os.path.exists(SDdraft):
    arcpy.Delete_management(SDdraft)
if os.path.exists(SD):
    arcpy.Delete_management(SD)

# Location of the project and map
aprx = arcpy.mp.ArcGISProject(wrkspc + "ToneDev.aprx")
aprxMap = aprx.listMaps("Map")[0]

arcpy.mp.CreateWebLayerSDDraft(aprxMap, SDdraft, serviceName,
'MY_HOSTED_SERVICES', 'FEATURE_ACCESS',
                              'Madrid', True, True, '', '', '', summary,
tags, '', '', '')

# Create service definition
arcpy.StageService_server(SDdraft, SD)

# connect to the portal
print("Connecting to {}".format(portal))
gis = GIS(portal, user, password)

# Find the SD, update it, publish /w overwrite and set sharing
sdItem = gis.content.search("{} AND owner:{}".format(serviceName, user),
                            item_type="Service Definition")[0]

sdItem.update(data=SD, thumbnail=thumbn)
fs = sdItem.publish(overwrite=True)

if shrOrg or shrEveryone or shrGroups:
  print("Setting sharing options...")
  fs.share(org=shrOrg, everyone=shrEveryone, groups=shrGroups)
print("Finished updating and sharing: {} - ID: {}".format(fs.title, fs.id))
```