Received November 16, 2016, accepted January 6, 2017, date of publication July 11, 2017, date of current version July 31, 2017.

*Digital Object Identifier 10.1109/ACCESS.2017.2725984*

# An Agent Based Traffic Regulation System for the Roadside Air Quality Control

**ABDELAZIZ EL FAZZIKI[1], DJAMAL BENSLIMANE[2], ABDERRAHMANE SADIQ[1], JAMAL OUARZAZI[3], AND MOHAMED SADGAL[1]**

[1]Computer Systems Engineering Laboratory, Cadi Ayyad University, 40000, Marrakesh, Morocco
[2]LIRIS Laboratory, Université Claude Bernard Lyon 1, 69622, Villeurbanne, France
[3]Laboratoire Physico-Chimie des Matériaux et Environnement (URAC 20), Cadi Ayyad University, 40000, Marrakesh, Morocco

Corresponding author: Djamal Benslimane (djamal.benslimane@univ-lyon1.fr)

**ABSTRACT** This paper describes an on-road air quality monitoring and control approach by proposing an agent-based system for modeling the urban road network infrastructure, establishing the real-time and predicted air pollution indexes in different road segments and generating recommendations and regulation proposals for road users. This can help by reducing vehicle emissions in the most polluted road sections, optimizing the pollution levels while maximizing the vehicle flow. For this, we use data sets gathered from a set of air quality monitoring stations, embedded low-cost e-participatory pollution sensors, contextual data, and the road network available data. These data are used in the air quality indexes calculation and then the generation of a dynamic traffic network. This network is represented by a weighted graph in which the edges weights evolve according to the pollution indexes. In this paper, we propose to combine the benefits of agent technology with both machine learning and big data tools. An artificial neural networks model and the Dijkstra algorithm are used for air quality prediction and the least polluted path finding in the road network. All data processing tasks are performed over a Hadoop-based framework: HBase and MapReduce.

**INDEX TERMS** Air quality management, crowd-sourcing, Dijkstra algorithm, mobile sensors, pollution prediction, traffic regulation.

## I. INTRODUCTION

Several studies have shown that the risks associated with cardiovascular and respiratory morbidity increase with chronic exposure to air pollution. Also, acute inhalation of pollutants even for a short period of time can lead to a dysfunction in the cardiovascular system and lung function [1], [2]. As in most cities, the basic pollution problem arises because the residential areas are far from workplaces, resulting in daily large population movements [3]. Thus, transport, especially road traffic, is a major source of air pollution in most of the cases. It contributes to 15% of the $CO_2$ emissions in Europe . In Moroccan urban environment, there is a range of pollutants in the atmosphere with the capacity to cause harm to both humans and the ecosystem, including Carbon monoxide, Nitrogen oxides, Sulphur dioxide, Particular matter, Volatile organic compounds, Ozone, and Hydrocarbons; thus, our particular interest in addressing this issue.

Therefore, many government agencies are increasingly interested in more monitoring and publication of data on air quality. At this level, two fundamental limitations on the approaches to the monitoring of air quality are imposed: First, the spatial and temporal resolution of pollution sampling is very low. For example, there exist about 3 active monitoring stations in Marrakech, separated from each other by tens of kilometers. This requires the use of mathematical models to calculate and estimate pollutant's levels in a large areas of the city, which can be both complex (requiring parameters such as the topography data, weather variables and chemical compositions) and inaccurate and therefore, can lead to incorrect conclusions [2], [4]. The estimation of personal exposure is essential for individuals to manage the risks associated with this problem, both through a retrospective understanding of the levels of pollutants that affect their health and the prospectively decision making to reduce the risks of being exposed to the pollution pikes [5].

Local authorities face a difficult situation with a constant increase in road traffic, leading to congestion and travel times that are constantly growing. This leads to an increase in fuel consumption and pollutant emissions, especially during peak hours. Emissions in cities often exceed the tolerated level; studies have shown that the effects of air pollution reduce life expectancy and induce enormous cost to society. However, the existing techniques for air quality and traffic congestion management have not been very effective because of the lack of autonomy, adaptation, and collaboration between the relevant entities. This issue needs the use of an integrated system for traffic management and air quality control, which aims to jointly optimize traffic and air quality. Therefore, an effective intelligent transport system taking into account air quality is of great importance for society.

In this work, we are particularly interested in traffic regulation and generating recommendations to take into account in order to improve air quality. The proposed system must be able to establish air quality indices in urban areas, to generate relevant information to users and propose recommendations about the best paths to take. This is done based on pollutant data collected from users embedded devices that provide properties related to road infrastructure and air quality to the system. It must take into account the constraints of geographical distribution and dynamics of the problem seen the fact that air quality levels change over time and the exact number of users which demand recommendations and then the number of vehicles to redirect to alternative paths is unknown. Therefore, the system must be adapted to user needs, taking into account all these factors while ensuring a great performance in handling the large amounts of data (pollutants, weather, etc.) gathered from different data sources (users devices, monitoring stations, weather data records, etc.) that will continually increase over time. For this, we propose the use of an approach based on agent technology [6], [7], crowd-sourcing, and Big Data analysis tools in order to measure, monitor and control the quality of air with a higher spatiotemporal resolution and to involve users in tracking their exposure to pollution through custom embedded tools.

Moreover, due to large amounts of data, the processing task has become a great challenge. To address this problem we suggest the use of Hadoop framework to ensure a great flexibility and speed in performing the needed prediction and analysis algorithms applicable to large scale data. Our main contributions in this work are:

1) Offer a low-cost component used by users who want to contribute to the monitoring of air quality by participating in the data collection task. This presupposes the development of a monitoring sensor unit to use onboard vehicles or by users, an application for data loading and transfer, and a server for data hosting.
2) A component for extracting and loading meteorological and pollution data from fixed monitoring stations.
3) Predicting the level of pollution near roads using the available data and ANN that have a strong ability to predict fuzzy data and efficiency in the modeling of dynamic systems.
4) Calculate the real-time and predicted air quality indexes, based on Murena method [8], and generating a spatio-temporal map for the distribution of pollution level in the study area.
5) The use of Dijkstra algorithm in the shortest path (least polluted path) finding [9], [10]. This algorithm has been widely used in route management systems, with variations to improve its performance [9].
6) We tested and deployed the system by simulating the data collecting by a small number of users to validate our proposal.

The rest of this paper is illustrated over a few sections starting with a brief literature review followed by an overview of the proposed system and the traffic regulation process details in section 3. Section 4 is devoted to the use of ANN for air quality prediction. The MapReduce based data analysis process is described in Section 5. In section 6 we present the used method for path finding in a road network. Section 7 is dedicated to the multi-agent system description. In section 8, a case study and the experimental results are presented, followed by a conclusion and perspectives in section 9.

## II. RELATED WORKS

In the past few years, many researches were centered around the use of static air quality monitoring stations along with crowdsourcing and participatory sensing. Such solutions lend themselves not only to monitoring the state of the physical world but can also help raising peoples awareness of issues related to air quality and pollution [2]. One of the main projects which proposed such a solution is the MESSAGE (Mobile Environmental Sensing System Across Grid Environments) project [11]. It aims to develop fixed and portable devices for high-density measurement of concentrations of carbon monoxide and nitrogen oxides in urban areas. They have very recently reported their development and deployment experience [12] in the Cambridge area, and demonstrated that the use of low-cost fixed and portable devices deployed in high densities can give a much more accurate picture of the spatial and temporal structure of air quality in the urban environment. In [13] authors introduced a low-cost mobile sensors based participatory air quality monitoring system. this work presents the design and implementation of GasMobile a small and portable measurement tool suited to be used by a large number of people.

In recent years, many works have been interested in using artificial intelligence and machine learning methodologies to address problems related to quantitative and qualitative measures more effectively. They can allow solving complex issues related to air quality management, traffic regulation and transportation systems by integrating intelligent components [14] Multi-agent based systems have been considered as an efficient tool for large-scale system such as intelligent traffic and air quality management [15]. They were known by the ability to model complex systems where a set of autonomous entities interact to produce global solutions [16]. Many works propose the use of the MAS technology in traffic control and management systems, such as [9] which proposes an integrated approach for modeling transport infrastructure and optimizing traffic in urban areas in order to reduce carbon-dioxide emissions. Authors in [14] have proposed an adaptive multi-agent system based on the ant colony behavior and the hierarchical fuzzy model. The solution is based on an adaptive vehicle route guidance system which allows adjusting efficiently the road traffic according to the real time changes in road networks. In most works based on MAS technology, such as [7], authors have integrated advanced analysis components and tools. Thus, the software agents use data mining methods for knowledge discovery, which will be used as a foundation for recommendation generation. Namoun *et al.* [9] have used an improved version of the Dijkstra algorithm as a graph search algorithm. This algorithm has a powerful potential to solve the problems of traffic congestion in a road network, which can be considered as a collection of graphs. The suggested solution in [9] combines the benefits of MAS and real time traffic data forecasting and analysis in the recommendation generation process. Machine learning tools such as ANN has been frequently used to predict air quality and pollution levels using a set of inputs, like pollutant concentrations, meteorological data and available traffic information [17], [18]. Moreover, many research works have been dedicated to implementing such computationally expensive algorithms and analysis tasks on parallel or distributed computing systems such as Hadoop [19]. Thus, information system enhancement using Hadoop as a data hub to optimize the information systems performance is considered as a new emerging strategy [20]. Many works such as [21] has proposed a study on big data integration with data warehouse built using relational technology mainly for operational sources. Regarding the presented related works, the development of management and control systems must be based on emerging technologies such as intelligence systems, Big Data and advanced analytics tools. It has been proved that MAS is a powerful computing paradigm to manage the complexity in dynamic distributed environment and the combination of MAS and Big data analytics tools can considerably improve the system performance.

## III. AIR QUALITY MANAGEMENT: TRAFFIC REGULATION

The aim of this work is to propose a hybrid solution to the issue related to the on-road air quality optimizing. This solution relies on both the data gathered from a set of static air quality monitoring stations and datasets gathered from several low-cost mobile sensors units. These units are embedded in users vehicles and connected to these users mobile devices (i.e. mobile phones) in order to upload the obtained data along with other additional data which we can get from these devices such as the GPS (Global Positioning System) location and temperature. The use of mobile sensors has the advantage of providing real time measurements from different locations (all the places visited by users) and also the contextual information that we can collect from the used mobile phones. Furthermore, the static air quality monitoring stations provide measurements datasets for various pollutants as well as meteorological parameters (Temperature, wind speed and direction, precipitation, pressure, etc) very accurately at a

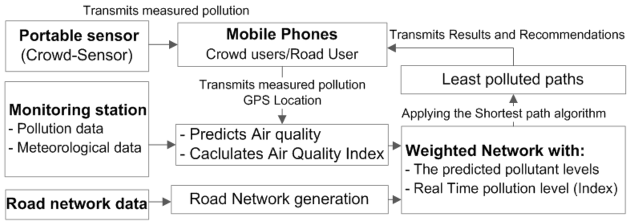



**FIGURE 1. Overview of the system structure.**

handful of fixed points across the study area (As illustrated in Figure 1).

One novel aspect of the system is that it makes use of immediate contextual parameters which can be retrieved from the road users' devices, such as time of day, location, speed and direction. These services help estimating traffic flow and congestion. As a result, users receive more relevant information and recommendations that are based on a combination of their historical preferences and contextual parameters.

All these types of inputs feed the system, which uses an advanced computational module for analyzing and extracting knowledge based on required algorithms and OLAP analysis. This module uses the pollutants data, contextual data and meteorological information for forecasting the pollutants concentrations in a specific zone by using an ANN based prediction model. The real-time and predicted results are then used in order to calculate an Atmospheric pollution index (API) based on Murena method (REF). The API is used to determine the quality of air in a simple manner and provides information that can be understood by the general public. The API is based on predicted rates of each pollutant comprising of sulfur dioxide ($SO_2$), ozone ($O_3$), nitrogen dioxide ($NO_2$), Atmospheric particulate matter ($PM_{10}$) and carbon monoxide (CO).

The computational module uses the resulting information in addition to the available road infrastructure data to build a weighted road network in which the edges weights are calculated based on the obtained APIs and the road segment lengths. The generated network is then used to find the most environment-friendly routes, provide forecasts about on-road pollution level and formulate recommendations to road users. The objective is to avoid the most polluted sections while averting the generation of traffic congestion. Figure 1 illustrates an overview of the system structuring.

As result, the main features of the system are:

- An effective communication with the users. It provides route solutions and recommendations that satisfy the preferences of users (e.g. Time of travel).
- Integrate and manage a wide range of data sources.
- Use of real-time data: receive real-time information about pollution levels, perform needed analysis and generate the air quality indexes for each urban path segment.
- Predicted air qualitys data: use historical data about pollutants and meteorological data to perform predictions and calculate the air quality indexes for path segments.

### A. POLLUTION MEASUREMENT SENSORS

One of the most critical and challenging component of the architecture is the device that will be used to measure air pollution levels. There are several options and we considered few criteria to choose the most suitable one:

- Portability: In order to face the issue related to the spatial coverage, the device must be made portable enough for a user to carry himself (e.g. on vehicles).
- Complexity: Having decided to make our devices suitable for vehicle-mounting, the next major decision we were confronted to was regarding target cost and complexity of the device.
- Sensor Type: depends on the pollutants to measure.

In this work we analyze the case where several users are equipped with ozone sensors along with a GPS module. These sensors are connected (via Bluetooth) with mobile devices which are equipped with a 4G communications unit. Mobile devices upload all gathered data to a central server which performs all required analysis and generates the suitable recommendations. Figure 2, illustrates the proposed process.

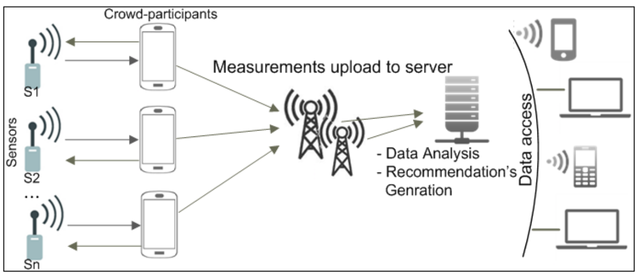



**FIGURE 2.** Data gathering and processing tasks.

### B. THE TRAFFIC REGULATION PROCESS

The traffic regulation and recommendation generation process is divided into the following steps (See Figure 3):

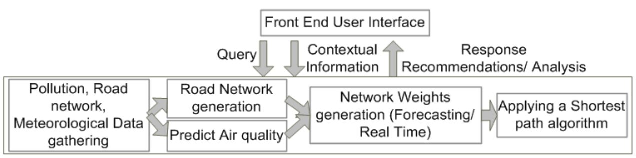



**FIGURE 3.** The traffic regulation and recommendation generation process.

- Pollution, meteorological and road network data gathering.
- Users' contextual information integration.
- Predict the air quality indexes.
- Road network generation.
- Data processing and the road network weights calculation.
- Apply the shortest path algorithm.

### C. ROAD NETWORK MODELING

A road network of the studied area is used in order to represents the transport supply. It is described by the different road segments and their intersections. In this work, we simulate the road segments for which properties are gathered from different data sources and provided to the system [22]. These collected data are used in the weighted graph generation (see Figure 4). A planar graph is used to formalize the

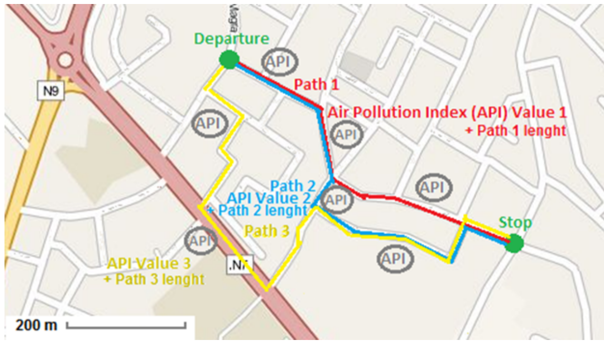



**FIGURE 4.** Road segment details and an example of road paths with different data (by Google Maps).

road network. In this graph,the roads are represented by a set of arcs and the junction points by nodes. Arcs are characterized by features such as APIs, path lengths and traffic capacity. Nodes are associated with intersections characteristics. The cartography is represented by a weighted graph $G = (V, E)$ where V is a set of vertices representing and $E = V \times V$ is a set of edges $e = (v_i, v_j)$. Each segment joining adjacent vertices is represented by either one or two directed edges. The edge weight $w_{ij}$, $v_j$ between two vertices $v_i$ and $v_j$ is a dynamic factor, which is used in order to reflect properties related to the edge $(v_i, v_j)$ which represents. Figure 5 illustrates an example of a road graph.

### D. ANN FOR AIR QUALITY PREDICTION

The main stage of air quality control is the process of predicting the on-road pollution level. The present study investigates the advantage of using neural networks for forecasting the on road air pollution. Artificial neural network (ANN) models are mathematical models inspired by the functioning of nervous systems, which are composed by a set of interconnected artificial neurons. These neurons can be associated in many different ways, depending on the characteristics of the issue to address. To construct an ANN model the air pollution system is considered as a system that receives information from distinct sets of inputs and produces a specific output [23].

The proposed network has been used for predicting pollutants concentration using three years data which contains hourly records of the different pollutants concentration and meteorological records. This prediction process is based on three stages. The first one is the data extraction stage, in which the objective is to choose the more significant data for the learning phase. The learning stage is the second phase. It aims to find the optimal configuration of hidden layers, the transfer

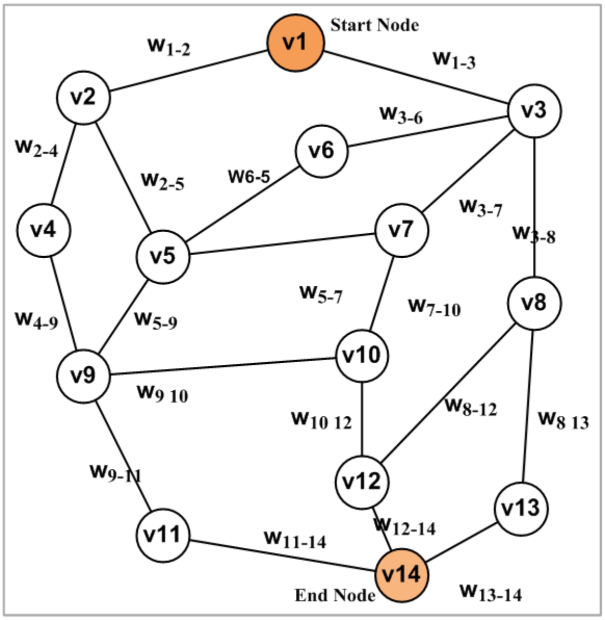



**FIGURE 5.** Example of a two directed edge based road graph.

function, and the performance index. The objective during this stage is to minimize the prediction error. The third stage is the prediction part in which we predict the pollutants concentration for a given time and location from a previously calibrated neural network during the learning phase.

In this work, we used static monitoring stations data which have hourly frequency in the prediction model. The available data records for each road segment is variable (from 4625 to 10124 records) due to the sensors deficient time. Each record contains 13 attribute: The five pollutants concentration ($SO_2$, $O_3$, $NO_2$, $PM_{10}$ and CO), solar radiation, Wind speed, Temperature (Celsius), Humidity ratio, Hour, Day, Month and year. The attributes and the hidden layers to use in the ANN model depend on the pollutant to predict. The network requires a data set to be calibrated. This set is called training set. The evaluation of the network is carried out using different data from those used during training, which is called the test set. The training set corresponds to the first 80% of records, and the tests set corresponds to the last 20%. We choose backpropagation as a learning rule since it is most suited and generally used technique for prediction problems. Along with, we choose to take advantage of the momentum to improve the efficiency of the algorithm. The evaluation of the neural network performance is carried through the Mean Square Error (MSE).

## IV. MAPREDUCE BASED DATA ANALYSIS PROCESS

The weighted graph generation task uses a multi-phase MapReduce process to get pollution level on different road segments [21]. The predicted and monitoring air pollution data (pollutants concentration) are loaded from the HBase in which they are stored. In the First Map stage (see Figure 6), we use the pollutants level datasets to calculate the pollution sub-indexes by applying the Murena [8] method. At the same time we use meteorological, geographical and road data files to generate the final API for each road segment [21], [24]. The intermediate results are stored into the output databases. In the Second Map phase (Figure 7) the API values are combined with users data and road network data in order to generate the weighted network. The intermediate results use the pair of segment identifier, and timestamp as the key map and the API as well as edges length as the value. During this Map stage the Dijkstra algorithm is applied in order to generate three shortest paths in response to each user query. In the Reduce phase, the paths with the same key are cumulated together.

## V. PATH FINDING IN A ROAD NETWORK

### A. WEIGHTED NETWORK GENERATION

The weights of the network segments are calculated by combining the predicted API records and the segment lengths. Murena method [8] was adopted to calculate the API. The evaluation of the API at stations for a pollutant p (PIs, p) is carried out by a linear interpolation between the reference scale values reported in Table 1 and is given by (Equation(1)):

$$PI_{s,p} = [\frac{PI_{hi} - PI_{lo}}{BP_{hi} - BP_{lo}}(C_p - BP_{lo}) + PI_{lo}]_{s,p} \qquad (1)$$

Where:

- $PI_{s,p}$:The value of the pollution index for a pollutant p at site s.
- $BP_{hi}$: The lowest break-point of a pollutant $p$ that is greater than or equal to $C_p$
- $BP_{lo}$: The highest break-point of a pollutant
- $p$ that is lower than or equal to $C_p$
- $PI_{hi}$: The $PI$ value corresponding to $BP_{hi}$
- $PI_{lo}$: The $PI$ value corresponding to $BP_{lo}$
- $C_p$: The pollutant $p$ daily concentration

An individual score is assigned to the level of each pollutant and the final API index is equal to the highest sub-index determined for each of the considered pollutant. The other required road segments properties are prepared and stored in internal databases. The weights of the network are updated when new forecasts are available. In this case, the system recalculates the API and dynamically updates the network with the new values. Otherwise the current weights are used to generate paths and recommendations.

### B. THE LEAST POLLUTED PATH SEARCHING: Dijkstra ALGORITHM

Dijkstra's algorithm is a graph search algorithm that solves the single source shortest path problem for a graph with nonnegative edge path costs, producing a shortest path tree. This algorithm is often used in routing and as a subroutine in other graph algorithms [25] In this work, we apply the Dijkstra shortest path finding algorithm over the generated traffic weighted networks in order to find the least polluted possible paths in the network. The weight of each segment is a combination of the different cost nature of using this

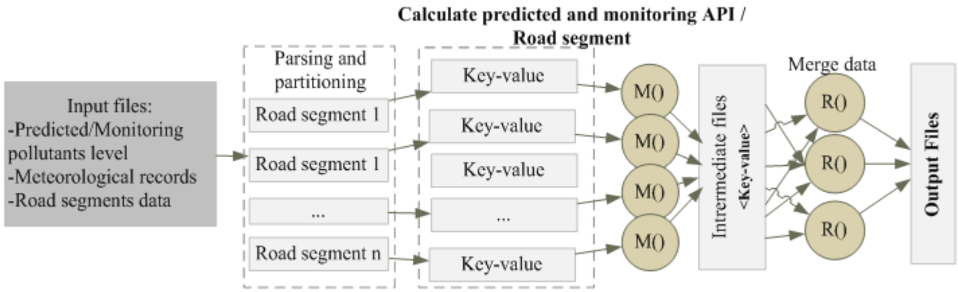



**FIGURE 6.** The first stage of the MapReduce process.

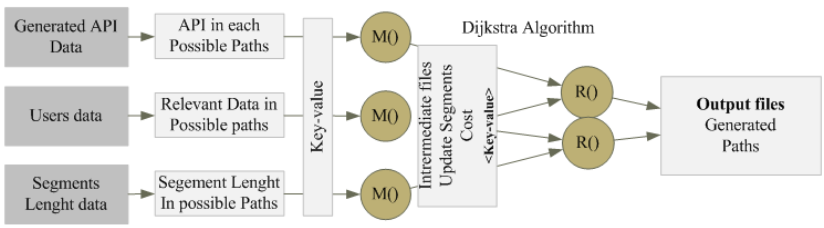



**FIGURE 7.** The second stage of the Map and Reduce process.

**TABLE 1.** Breakpoints for the proposed API ($\mu g/m^3$ for all pollutants and $mg/m^3$ for CO)(Murena 2004 [8]).

| Pollution level | API | PM10 | NO2 | CO | SO2 | O3 |
|---|---|---|---|---|---|---|
| Unhealthy | 85-100 | 238-500 | 950-1900 | 15,5-30 | 500-1000 | 223-500 |
| Unhealthy for sensitive groups | 70-85 | 144-238 | 400-950 | 11,6-15,5 | 250-500 | 180-223 |
| Moderate pollution | 50-70 | 50-144 | 200-400 | 10-11,6 | 125-250 | 120-180 |
| Low pollution | 25-50 | 20-50 | 40-200 | 4-10 | 20-125 | 65-120 |
| Good quality | 0-25 | 0-20 | 0-40 | 0-4 | 0-20 | 0-65 |

particular segment. As we are interested in the air quality optimizing by traffic regulation, we use the on-road calculated API values and the segment lengths [9].

By using this weighted network the system can process user queries, generates recommendations and calculate potential roads according to the Dijkstra shortest path algorithm. For a particular source node, the algorithm finds the path with the lowest cost between that node and any other node. The used function for calculating the shortest path will depend on the road segment length and the calculated API for each path segment. Thus, the value of the path can be provided by the following formula (Equation(2) and Equation(3)):

$$P_i = \sum_{i=0}^{n} X_i \times W_i \tag{2}$$

$$W_i = (C_1 \times A_i + C_2 \times L_i)/(C_1 + C_2) \tag{3}$$

Where, $X_i$ is a configurable weight and $W_i$ is the weight of the edge number $i$ and $n$ is the potential number of edges to browse. $A_i$ and $L_i$ are successively the current API and length of the edge i. $C_1$ and $C_2$ are two coefficients related to each cost parameter. They are controlled by decision makers in order to optimize the weight calculation function. In this work we modify the Dijkstra's shortest path algorithm to obtain the three selected shortest paths in order to provide the decision making and risk evaluation to the users.

## VI. THE MULTI-AGENT SYSTEM ARCHITECTURES

In order to provide an adequate solution in terms of robustness and agility, we use a multi-agent framework to represent the on-road air quality management system. The objective is to propose an architecture that consists of a set of autonomous agents able to set their own goals and actions and interact and collaborate with each other through a communication protocol [26]. The agents can be categorized into three groups: 1). A data integration agent group which continuously monitorsing and collectsing air pollution data, meteorological data, etc, 2). The agents in the data processing group analyze the pollution data provided to generate recommendations for road users, and 3). While the interface group agents which are responsible for contextual data gathering.

A two-level hierarchical MAS can also be adopted. The data processing group is cognitive agents that give instructions while the worker agents in UI and data integration groups are reactive agents that carry out the tasks assigned. Figure 8 illustrates the multi-agent system structure. In the

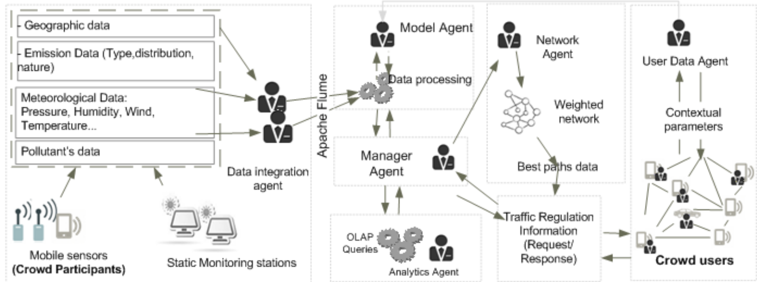



**FIGURE 8.** **The Multi-agent air quality system for traffic regulation.**

development process we chose to use the Prometheus methodology to design and implement the multi-agent system. Our choice is motivated by the efficiency of message communication and lightweight nature of the framework. Prometheus methodology [27] has been developed to support the complete software development life-cycle from problem description to implementation. It offers an environment for analyzing, designing, and developing heterogeneous multi-agent systems. This methodology consists of three phases: System Specification, architectural design and the detailed design. The Prometheus Development tool is extended with the ability to generate skeleton code in the JACK agent-oriented programming language [28] using a PDT code generator extension which maintains also synchronization between the generated code and the design when either of them changes.

The derived multi-agent system is composed of seven main agents:

- Data integration agents: These agents represent all monitoring stations and mobile sensors distributed in the study area and provide the required functionality during the data extraction, transformation and loading process. A data integration agent is also responsible for data validation, accuracy, the type conversion, etc. It prepares all input data required for the good functioning of the model agent. It sets up a register of emissions for the region in order to make regional modeling and prepare the needed domain data and meteorological parameters. Collaboration between these agents will allow a better understanding of the spatiotemporal evolution of surface air quality.
- Model agent: The Model agent performs the deterministic modeling. It brings together a set of equations representing the transport and chemistry of gaseous species, allowing the quantification of the evolution of a set of pollutants, according to time on road segments, taking into account all parameters. This agent is responsible for the generation of the air qualitys prediction using ANN, which is capable of modeling highly nonlinear relationships while taking into account the data distribution factors. This agent uses the predicted and real time data in order to calculate API values which are then used to generate the weighted network. All resulting data are stored into an HBase.
- Network agent: This agent is responsible of the road paths network generation and path finding in this weighted network by searching for the shortest possible route between two road nodes. The path is represented by a set of segments to browse from a source node to the destination node.
- Managing agent: Responsible for the reliability of the whole system, and manages the operation of the individual agents, especially the station agents, model, data provider and network generation.
- User's Data agent: This agent uses the mobile app to gather data concerning traffic, meteorological and road conditions using users device sensors readings (e.g. Accelerometer, GPS).
- Analytics agent: The purpose of the OLAP agent is to convert the amount of monitoring data into valuable information by applying quick and effective analysis and create recommendations for users using various views and representations of data. It provides all the basic functionality of an OLAP system and also the missing intelligence in traditional OLAP systems. The aim is performing OLAP analysis on behalf of an agent or a user and reporting its result back to the requesting entity and all other entities that should be informed [29].

## VII. CASE STUDY

### A. THE STUDY AREA

The study area used for the testing scenario in this work is located in Marrakech-City. This study area suffers from the effects of pollutants produced by vehicle exhaust systems. This study is based on a set of sensors that provide information and measures of the air pollutant concentration. It focused on air quality indexes related to the following pollutants: Sulfur Dioxide ($SO_2$), Nitrogen dioxides ($NO_2$), Carbon Monoxide (CO), Particulate Matter, and Ozone ($O_2$). The map used for the testing scenario in this paper was cut in order to reduce the simulation time. The area considered in the simulations is represented in Figure 9.

### B. AIR QUALITY PREDICTION

We use a three-layer perceptron ANN model and data concerning the study area described above to predict pollutants level. In this case study we generate prediction of the Ozone concentration; A six neurons in the input layer are

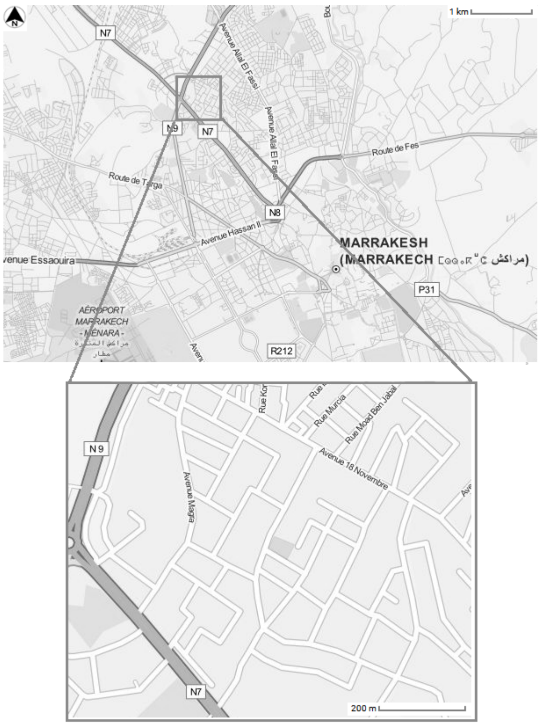


FIGURE 9. Inset map of an area of Marrakech used for simulations (by Google Maps).

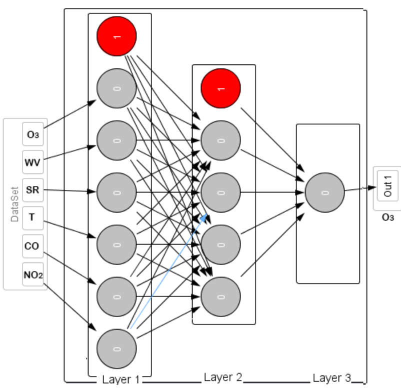


FIGURE 10. The three-layer ANN perception configuration.

used including temperature, solar radiation, the $NO_2$ concentration, CO concentration and wind velocity. The number of hidden layers and neurons in each hidden layer are the parameters to choose in the model construction. In our case, we tested different configuration and we found that a model based on one hidden layer and four neurons gives the best ANN performance. The last layer is the output, which consists of the target of the prediction model. Here, $O_3$ was used as the output variable and a hyperbolic tangent sigmoid function was used as the transfer function. A two year data set was divided into two parts: 80% used for training the networks and the remaining 20% employed in testing the networks. The mean square error was chosen as the statistical criteria for measuring the network performance. Figure 10 illustrates this ANN configuration. There is no exact procedure to meet the optimal number of hidden layers and neurons, or the value of momentum and the learning rate. We experimented multiple configurations until convergence of the MSE, in a way to find the optimal one. We concluded that more neurons do not help getting lower mean square error. In addition, the momentum and the learning rates must be low to get the best results.

The following graph (Figure 11) shows the performance of the network above. It represents a comparison of the observed and the predicted Ozone concentrations evolution based on the MSE according to time (during two days).


FIGURE 11. Comparison of one day observed and predicted Ozone concentration values from the ANN for one road segment.

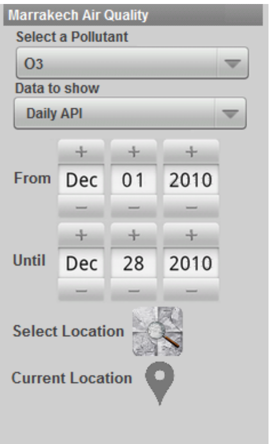



FIGURE 12. A screenshot from the used mobile user interface.

### C. API VALUES GENERATION

The system generates the analysis queries based on selected options from an easy to use user interface to get the resulting API. Figure 12 presents a screenshot of the systems final user interface which allows users to select the desired options from a web form and Figure 13 is an example of a query result which allows selecting the API during one week (On May 2015) in a selected location from the study area.

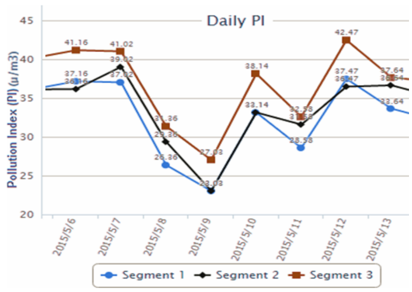



**FIGURE 13.** Three segments pollution index in the simulation area during one week.

### D. THE LEAST POLLUTED PATH FINDING

In the implemented system prototype, the road graph is generated in a specific location in the study area according to users' needs. For each segment in this location the system calculates or updates the costs which are calculated based on the available data and the Dijkstra algorithm is then applied. Figure 14 describes a road network example for a user who wants to navigate from the node N1 to N14 and the generated shortest path.

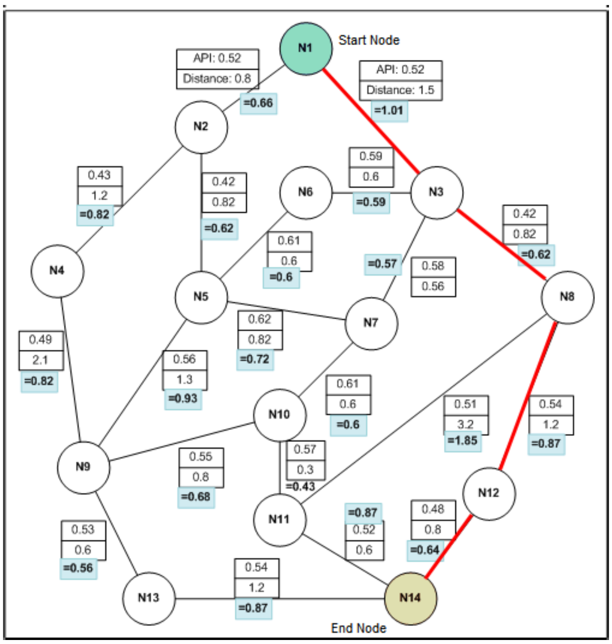



**FIGURE 14.** Example of a weighted road graph.

## VIII. CONCLUSION

The main contribution of this work is the definition of a development process based on big data and intelligent systems concepts for a traffic regulation system according to air quality data. We have, through this paper presented the implementation of an air quality system for recommendation and traffic regulation over distributed data gathered from different air quality sensors, users' devices and other external databases, that are managed using Hadoop to ensure fast data loading, fast query processing and efficient storage. The system is currently being tested using real-time Marrakech city data. The case study addresses the generation of pollutants API in the road segments. The pollution level is then used in order to calculate the best route a user can use.

Our experimental results show that the data processing operations and the algorithms deployed in the large-scale data processing system are feasible and efficient. The use of a system based on Hadoop has improved the performance and decreased significantly the processing time. The perspectives of this work are the integration of multi-criteria decision support tools and the use of historical traffic data in the decision making process in order to generate more appropriate recommendations for final users.

## ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of other members of computer systems engineering laboratory and physical-chemistry of materials and environment laboratory (URAC 20) for their helpful discussions and the availability of all data and resources that have helped make this work in the best conditions.

## REFERENCES

[1] N. L. Mills *et al.*, "Adverse cardiovascular effects of air pollution," *Nat. Clin. Pract. Cardiovascular Med.*, vol. 6, no. 1, pp. 36–44, 2009.

[2] V. Sivaraman, J. Carrapetta, K. Hu, and B. G. Luxan, "HazeWatch: A participatory sensor system for monitoring air pollution in Sydney," in *Proc. 38th Annu. IEEE Conf. Local Comput. Netw. Workshops*, Oct. 2013, pp. 56–64. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6758498

[3] EUROPEAN COMMISSION and Climate Action. *Road Transport: Reducing $CO_2$ Emissions From Vehicles*, accessed on Nov. 11, 2016. [Online]. Available: http://ec.europa.eu/clima/policies/transport/vehicles/index en.htm

[4] L. J. Young, C. A. Gotway, J. Yang, G. Kearney, and C. DuClos, "Linking health and environmental data in geographical analysis: It's so much more than centroids," *Spatial Spatio-Temporal Epidemiol.*, vol. 1, no. 1, pp. 73–84, 2009.

[5] D. Olaru and J. Powell, "What activity-based analysis and personal sampling can do for assessments of exposure to air pollutants?" in *Air Pollution Modeling and Its Application XIX*. Dordrecht, The Netherlands: Springer, 2008, pp. 717–718.

[6] I. T. Ćirić, Ž. M. Ćojbašić, V. D. Nikolić, P. M. Živković, and M. A. Tomić, "Air quality estimation by computational intelligence methodologies," *Thermal Sci.*, vol. 16, no. 2, pp. S493–S504, 2013. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84874808919&partnerID=40&md5=9af13791ba66850493af418a57387bff

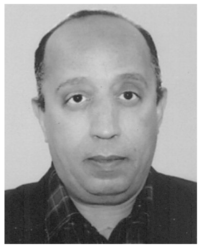
[7] M. V. Sokolova and A. Fernández-Caballero, "Modeling and implementing an agent-based environmental health impact decision support system," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2603–2614, 2009.

[8] F. Murena, "Measuring air quality over large urban areas: Development and application of an air pollution index at the urban area of Naples," *Atmos. Environm.*, vol. 38, no. 36, pp. 6195–6202, 2004.

[9] A. Namoun, C. A. Marín, B. S. Germain, N. Mehandjiev, and J. Philips, "A multi-agent system for modelling urban transport infrastructure using intelligent traffic forecasts," in *Proc. 6th Int. Conf. (HoloMAS)*, vol. 8062, 2013, pp. 175–186. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84884969849&partnerID=40&md5=0a85fc95883c3c7911c22956260c410c

[10] H. Zahmatkesh, M. Saber, and M. Malekpour, "A new method for urban travel rout planning based on air pollution sensor data," *Current World Environ.*, vol. 10, pp. 699–704, May 2015. [Online]. Available: http://www.cwejournal.org/vol10nospl-issue-may-2015/a-new-method-for-urban-travel-rout-planning-based-on-air-pollution-sensor-data/

[11] R. Anderson and J. Giddings, "Message: A mobile environmental sensing system across grid environments," Dept. Comput., Imperial College London, London, U.K.: Tech. Rep., 2009.
[12] M. I. Mead *et al.*, "The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks," *Atmos. Environ.*, vol. 70, pp. 186–203, May 2013.
[13] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, "Participatory air pollution monitoring using smartphones," in *Proc. 2nd Int. Workshop Mobile Sens.*, 2012, pp. 1–5. [Online]. Available: ftp://ftp.tik.ee.ethz.ch/pub/people/hdavid/HSST2012.pdf
[14] H. M. Kammoun, I. Kallel, J. Casillas, A. Abraham, and A. M. Alimi, "Adapt-Traf: An adaptive multiagent road traffic management system based on hybrid ant-hierarchical fuzzy model," *Transp. Res. C, Emerg. Technol.*, vol. 42, pp. 147–167, May 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X14000692
[15] X. Jin and L. Jie, "A study of multi-agent based model for urban intelligent transport systems," *Int. J. Adv. Comput. Technol.*, vol. 4, pp. 126–134, Apr. 2012.
[16] G. Phillips-Wren and L. Jain, "Recent advances in intelligent decision technologies," in *Knowledge-Based Intelligent Information and Engineering Systems* (Lecture Notes in Computer Science), vol. 4692. Berlin, Germany: Springer, 2007, pp. 567–571.
[17] T. Fontes, L. M. Silva, S. R. Pereira, and M. C. Coelho, "Application of artificial neural networks to predict the impact of traffic emissions on human health," in *Progress in Artificial Intelligence* (Lecture Notes in Computer Science), vol. 8154. Berlin, Germany: Springer, 2006, pp. 21–29.
[18] I. García, J. G. Rodriguez, and Y. M. Tenorio, "Artificial neural network models for prediction of ozone concentrations in Guadalajara, Mexico," in *Air Quality-Models and Applications*. Rijeka, Croatia: InTech, Jun. 2011. [Online]. Available: http://www.intechopen.com/books/air-quality-models-and-applications/artificial-neural-network-models-for-prediction-of-ozone-concentrations-in-guadalajara-mexico
[19] Apache.Org. *Apache Hadoop Documentation*. [Online]. Available: http://hadoop.apache.org
[20] T. Das and A. Mohapatro, "A study on big data integration with data warehouse," *Int. J. Comput. Trends Technol.*, vol. 9, no. 4, pp. 188–192, Mar. 2014. [Online]. Available: https://doaj.org/article/3d9781bcab794e26b11f44ccefbc97fc
[21] J. Zhao, J. Zhang, S. Jia, Q. Li, and Y. Zhu, "A MapReduce framework for on-road mobile fossil fuel combustion $CO_2$ emission estimation," in *Proc. 19th Int. Conf. Geoinformat.*, Jun. 2011, pp. 1–4. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=5980759
[22] G. Mitran, "The estimation of air pollution from road traffic by transport modeling," *Int. J. Sci., Techn. Innov. Ind. Mach., Technol.*, vol. 5, no. 4, pp. 60–63, 2007.
[23] A. Russo, F. Raischel, and P. G. Lind, "Air quality prediction using optimal neural networks with stochastic variables," *Atmos. Environm.*, vol. 79, pp. 822–830, Nov. 2013.
[24] W. Fang, V. S. Sheng, X. Wen, and W. Pan, "Meteorological data analysis using MapReduce," *Sci. World J.*, vol. 2014, pp. 1–10, Feb. 2014.
[25] D. Fan and P. Shi, "Improvement of Dijkstra's algorithm and its application in route planning," in *Proc. 7th Int. Conf. Fuzzy Syst. Knowl. Discovery (FSKD)*, Yantai, China, 2010, pp. 1901–1904.
[26] D. Lavbič and R. Rupnik, "Multi-agent system for decision support in enterprises," *J. Inf. Org. Sci.*, vol. 33, no. 2, pp. 269–284, 2009.
[27] P. Lin, J. Thangarajah, and M. Winikoff, "AUML protocols and code generation in the Prometheus design tool," in *Proc. 6th Int. Joint Conf. Auto. Agents Multiagent Syst. (AAMAS)*, Honolulu, HI, USA, 2007, p. 270.
[28] P. Busetta, R. Rönnquist, A. Hodgson, and A. Lucas, "JACK Intelligent Agents-Components for Intelligent Agents in Java," *AgentLink News*, vol. 1, no. 2, pp. 2–5, 1999.
[29] S. Muhammad, "Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool," Ph.D. dissertation, Lund Univ., Lund, Sweden, 2010. [Online]. Available: http://lup.lub.lu.se/student-papers/record/3559141/file/3559170.pdf



**ABDELAZIZ EL FAZZIKI** received the M.S. degree from the University of Nancy, France, in 1985, and the Ph.D. degree in computer science from Cadi Ayyad University in 2002.

He has been with Cadi Ayyad University since 1985, where he is currently a Professor of computer science. He has been responsible for the master's degree program in information system engineering since 2006.

He was the Director of the Engineering Computer Systems Laboratory between 2011 and 2015. He has co-authored several papers on agent-based image processing, and is the main Author of over 20 papers in software engineering and data analytics field. His research interests are related to software engineering, decision support, big data, data analytics, crowdsourcing, and e-government. In the MDA field, he has been involved in agent-based systems, service-oriented systems, and decisional systems.

**DJAMAL BENSLIMANE** is currently a Full Professor of Computer Sciences with Université Claude Bernard Lyon 1 and a member of the Lyon Research Center for Images and Information Systems, Service Oriented Computing Research Team, Lyon, France.

He has authored papers in well-known journals, including the IEEE TKDE, ACM TOIT, SIGMOD Record, the IEEE TSC, the IEEE INTERNET COMPUTING, the *WWW Journal*, and the *DPDB Journal*. His research interests include distributed information systems, Web services, ontologies, and databases.

**ABDERRAHMANE SADIQ** received the M.S. degree in 2012 and the Ph.D. degree in computer science from the University of Marrakesh in 2017.

He has authored or co-authored several papers on software engineering, data analytics, and air quality management. His research interests are related to software engineering, decision support, data analytics, crowdsourcing, and air quality management.

**JAMAL OUARZAZI** is currently a Full Professor of Chemistry with Cadi Ayyad University and a member of the Laboratory of Physical chemistry of Materials and Environment, Marrakesh, Morocco.

He has authored papers in well-known journals, including *Chemical Engineering*, *Inorganic Chemistry*, *Thermochimica Acta*, *Magnetic Resonance in Chemistry*, *Inorganica Chimica Acta*, and *SpringerPlus*. His research interests include statistical and deterministic modeling of the air quality.

**MOHAMED SADGAL** received the Ph.D. degree in computer science from the University of Lyon in 1989, and the Ph.D. degree in computer science from Cadi Ayyad University in 2005. From 1985 to 1987, he was an Associate Researcher with Lyon I, France. He is currently a Professor with Cadi Ayyad University, Marrakesh, Morocco. His research interests include computer vision, artificial intelligence, and multi-agent systems.