Analyzing Air Pollution on the Urban Environment

Analyzing air pollution on the urban environment

Elena Baralis, Tania Cerquitelli, Silvia Chiusano, Paolo Garza, and Mohammad Reza Kavoosifar
Department of Control and Computer Engineering, Politecnico di Torino, Torino, Italy
{name.surname}@polito.it

Abstract - Air pollution is one of the most important factor that can affect the quality of citizen life in the urban environment. Consequently, monitoring air pollution is currently a critical issue that needs to be addressed for enhancing the well being of citizens. This paper proposes a data analysis engine, based on business intelligence methodologies and open technologies, to support different targeted analysis on air pollution data. To analyse the problem from different facets, air pollution measurements are enriched with additional information such as meteorological and traffic data, which are collected through sensor networks available in the smart city context. This integrated dataset is periodically analyzed to generate informative dashboards based on a selection of Key Performance Indicators (KPIs). The informative dashboards can provide useful insights about pollutant concentration at different time granularity levels in the urban areas and support a joint evaluation of pollutant concentrations with climate conditions and traffic flow. As a reference use case, open data on air pollution in the urban area of a major Italian city is analyzed to demonstrate the effectiveness of the proposed approach in a real smart city context.

I. INTRODUCTION

Today's there is an increasing concern of citizens and city administrations on air pollution, because of its possible significant impact on human health. The analysis of air pollution involves the simultaneous study of additional data that may impact on air quality such as traffic flow and meteorology. Consequently, the abundance of information available from different sensor networks deployed in the urban city provides an unprecedented opportunity to understand the quality of life of city dwellers and how it deteriorates over time due to air pollution.

This paper presents the APA (Air Pollution Analysis) system, based on business intelligence methodologies and open technologies, to efficiently support different targeted analyses for increasing user awareness on air quality in the city environment. Data on air pollution has been integrated with weather and traffic data to build a richer data collection and effectively support (i) the evaluation of the *pollutant concentration* in different time periods and urban areas, the analysis of the impact of (ii) *climate conditions* and (iii) *traffic of vehicles* on air pollution.

We defined a set of Key Performance Indicators (KPIs) to monitor the environmental pollution together with weather and traffic data at different spatial and temporal granularity levels. For example, a KPI has been defined to evaluate the concentration of a given air pollutant in a time period, its variation over time, or how many times it was above a given reference threshold in a given time period and spatial area. KPI values are graphically visualized (also on geographical maps) on informative dashboards to support an easy understanding of the results of the analysis and thus enhance user awareness on air quality. This analysis can provide useful information for the definition of medium- and long-term intervention policies aimed at reducing the concentration of pollutants. For example, the proposed informative dashboards can provide useful feedbacks about *when* and *where* pollutants reached critical concentrations. This information may support the local Administration in planning/re-planning the paths used for the distribution of goods in some urban zones or sizing alternative infrastructures to support (green) urban mobility (such as public transport and bike sharing systems) in some parts of the city.

Although the APA system is general and it can be applied to data acquired in different city environments, to assess the proposed approach we consider as use case scenario the analysis of air pollution data monitored in a major Italian city.

This paper is organized as follows. Section II discusses the main research activities addressing the analysis of air pollution. Section IIII describes the main building blocks of the proposed business intelligence system. Section IV summarizes the development and experiments on real data. Section V draws conclusions and presents future developments of this work.

II. RELATED WORK

Today's world cities are negatively affected by pollution that significantly deteriorates the quality of life of dwellers. Air quality can vary over time and across different areas of a city, and it is also influenced by different factors such as weather conditions (e.g., humidity, temperature and atmospheric pressure), human activities (e.g., traffic flows, people's mobility), provided services and the presence of points of interest in urban areas [1]. Monitoring air quality, in terms of pollutant concentrations such as PM2.5, NO2, and SO2 is usually performed by means of a set of fixed stations deployed in the urban environment. Since the costs of designing, developing, deploying and maintaining these stations are high, they are usually deployed in a limited number.

With the evolution of mobile communication protocols and sensing technologies innovative sensors have been developed to monitor a wide range of pollutant, thus evaluating air quality. Authors in [2] proposed to monitor the air quality through the deployment of mobile sensors on the bicycle wheels. Such sensors are able to monitor CO2 and some meteorological data such as temperature and atmospheric pressure. Despite their high potential, these approaches allow monitoring a limited number of gaseous substances (such as CO2 and CO), while other pollutants (such as PM2.5 and PM10) require devices not easily portable and long periods of monitoring to generate accurate measurements [1].

A parallel effort has been devoted to designing innovative business intelligence and/or data mining solutions to perform different targeted and interesting analyses on pollutant measurements to evaluate air quality. Authors in [3] studied the pollutant concentration in different cities, or in different areas of a city, its variation over time and the correlation degree between concentrations of pollutants and other information such as weather conditions. Pollutant measurements were collected through a network of fixed monitoring stations, integrated with meteorological data and stored in a data warehouse. To monitor air quality different indicators were defined to perform the analyses at a different spatial-temporal granularity. The APA engine addresses the research issue discussed in [3]. However, APA exploits different technological solutions and supports a richer set of analyses because traffic data are also integrated in the system.

Data Mining techniques were used to support the analysis and the prediction of air quality in the urban environment [1, 4]. Authors in [1] proposed to exploit classification techniques to forecast the air quality level in areas without monitoring stations. Historical and real-time measurements on air quality together with additional information such as weather, traffic, and people's mobility have been analyzed as training data. Authors in [4] proposed a jointly analysis of historical data on air quality and weather forecasts to predict future values (in the next time periods) on air quality. In the current implementation of APA the business intelligence methodologies have been only integrated. However, data mining algorithms will enrich the second release of APA to address advanced air quality data analytics.

III. THE APA PLATFORM

The APA engine is designed for the modeling, integration, storage, and analysis of a large amount of heterogeneous data related to air pollution to provide different levels of relevant knowledge on air quality on their urban environment to users. APA supports different targeted analyses for different users, i.e., citizens and staff of the public administration.

To address air quality analyses, different data types provided by ad-hoc sensor networks deployed in the urban areas are integrated in APA. Specifically, data on *pollutant concentrations* were integrated with *climate data* and *traffic data* related to traffic of vehicles. All preprocessed data are stored into a unique data repository (a *data warehouse*). Different *key performance indicators* (KPIs) have been defined to monitor the concentration of various pollutants and to analyze their trend over time together with meteorological and traffic data. APA also includes *informative dashboards* to present the results of the analysis (e.g., trend of KPIs over time, correlations between two KPIs) on air quality in an informative fashion, using simple and easily understandable data representations. Each component of APA is detailed in the next sections.

A. Data sources

The APA engine is currently integrating three different types of data. In addition to measurements on pollutant concentration, APA also considers climate data and data on traffic of vehicles, since both aspects may have a significant impact on the air quality value. The considered data categories are briefly described below.

(i) Measurements on *pollutant concentrations* in the urban areas. Different pollutant are currently analysed in the APA engine including particulate matter (PM10 and PM2.5), carbon monoxide (CO), ozone (O3), nitrogen dioxide (NO2), benzene (C6H6), sulfur dioxide (SO2).

(ii) *Climate conditions* in the urban areas, including air temperature, relative humidity, precipitation level, wind speed and atmospheric pressure.

(iii) *Vehicle traffic data* as the number of vehicles entering in a given urban area at a given time granularity (e.g., hourly). To characterize traffic under different perspectives, we collected traffic data for different categories of vehicles. Specifically, we categorized vehicles based on their fuel type (e.g., petrol, diesel) and their use (e.g., bus, private vehicles or transport of goods).

For each of the above data types, measurements in the urban environment are usually collected at a given time granularity through ad-hoc *geo-referenced sensor networks* deployed in the city area.

Concentration measurements for each pollutant were collected through sensors deployed in monitoring stations located in different areas of the city. Each *pollution monitoring station* (PolMS) is characterized by the geo-coordinates (i.e., latitude and longitude) of its location. It includes different sensors to monitor the concentrations of various pollutants at hourly and daily time granularity.

For monitoring climate conditions, the Weather Underground web service[1] has been considered. It gathers data from a geo-referenced network of Personal Weather Stations (PWS) registered by users. For many cities a large number of stations are distributed throughout the territory. Although the measurement frequency can be easily set by the user for each PWS (and can vary over time), the average value for the ones we considered was about 15 minutes.

Traffic data were collected through geo-referenced sensor networks measuring the number of vehicles (for each of the above considered vehicle types) entering, with hourly time granularity, in different urban areas.

B. Data preparation and modeling

To efficiently support different targeted analyses, data on pollutant concentrations are integrated with climate and traffic data. These data are additional measurements describing the environmental context in which the pollutant concentrations were monitored.

Available climate and traffic data are pre-processed before the data integration phase because of their (possible) different time and space granularities with respect to the pollutant timeline and the monitored areas. Since we are focusing on pollution analysis, the spatial-temporal granularity of the sensor network monitoring pollutant concentrations has been considered as a reference. Then, data coming from climate and traffic monitoring networks have been integrated using the same spatial-temporal granularity of the reference network.

For the weather Underground web service, a large number of Personal Weather Stations (PWS) are distributed throughout the territory for many cities, and the measurement frequency can be easily set by the user for each PWS. However, sensors networks monitoring pollutants and climate data may have different geo-locations and sampling rate in collecting measurements. For example, weather data may be unavailable for a specific pollution monitoring station (PolMS) located in a given urban area or at a given instant of time, while the pollutant measurements monitored in the station are instead available. To deal with these issues, in the data integration phase, weather data associated with a given pollution station are computed as a distance-based weighted mean of the values provided by the three nearest PWSs. The weight is inversely proportional to the distance from the three PWSs to the PolMS. Hence, three equally distant PWSs would have the same weight in determining the weather values of a given city zone. Weather data timestamps were aligned to the closest timestamp available for the given pollution monitoring station (PolMS) through an approximate join.

For traffic data readings, the number of entering vehicles in each area has been associated to all the sensors deployed in the area. Traffic data were timely integrated through an approximate join similar to that adopted for climate data integration.

Air quality data together with weather and traffic data are then stored into a unique data repository (a *data warehouse*), whose conceptual representation, according to the Dimensional Fact Model (DFM) [5], is reported Figure 1. The *fact table* consists of a main measure, i.e., the hourly pollutant concentration per sensor, and some additional measures describing the context in which the pollutant concentration was monitored. These additional measures are the climate conditions (e.g., hourly temperature, humidity, and wind speed per sensor) coming from outdoor PWSs, and traffic data per sensor (e.g., number of entering vehicles for each type of vehicle). Three *dimension hierarchies* are defined to analyse the spatial and temporal distribution of the air quality and the other contextual information with different granularity levels. Specifically, we defined a *location-related dimension hierarchy* linked to the sensor device monitoring the concentration of a given pollutant, and *two temporal-related dimension hierarchies* linked to the time for pollutant concentration values reported in the fact table. Dimensions are detailed below.

The *temporal-related dimension hierarchies* provide many different blends of time spans. In one hierarchy, time spans from hours to time slots. The hour is aggregated into different intervals (2-hours, 8-hours), and the corresponding daily time slot (morning, afternoon, evening, or night). In the other hierarchy, time spans from dates to years. Moreover, each date is classified as working or high day and as week day (e.g., Monday or Tuesday).

The *location-based dimension hierarchy* starts from physical sensors and builds up to the whole city area, with the type of sensor (i.e., the measured pollutant) as related feature included in the dimension. To analyze the spatial distribution of the air quality, higher-level space granularities are also considered beyond the geographical coordinate of the sensor from which the measurements refer to. Each sensor is mapped to the corresponding address and city area. While the geographical coordinate is recorded for the considered sensor, the address and the area name are added as additional contextual features to perform interesting analysis.

[1] http://www.wunderground.com/

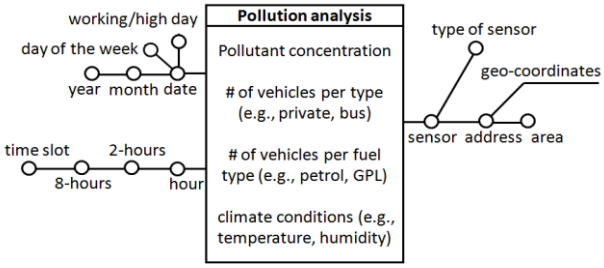

Figure 1 - Data warehouse conceptual model

C. Data analysis

The prepared data are analyzed to gain insights into the air pollution condition on the urban area. The aim of the analysis is to produce useful feedbacks to the end-users by generating informative dashboards using Key performance Indicators (KPIs).

We identified two operational roles representing users of the APA engine: (i) *staff of the local administration* and (ii) *citizens* of the urban environment. (i) The *staff of the local administration* is mainly interested in understanding the main causes of pollution affecting the city with the aim of guaranteeing a good level of air quality. These users are interested in monitoring air pollution in the urban area to point out *when* and *where* the pollutant concentrations became critical. They are also interested in assessing contextual aspects that may contribute to air pollution as the climate condition and the traffic of vehicles in the urban areas. It follows that these users need to analyze the complete streams of collected data (also at different spatial and temporal granularities), to observe and understand the observed phenomenon, assess the different components and identify possible causes. (ii) *Citizens* are interested in locally assessing the air quality in some specific urban areas, as for example where they live or work. These users need the visualization of a few indicators at different spatial and temporal granularities, but possibly provided in an easily understandable and intuitive way.

APA includes informative dashboards to present the results of the analysis (e.g., trend of KPIs over time, correlations between two KPIs) on air quality in an informative fashion, using simple and easily understandable comparisons among KPI values. In our context, KPIs are quantitative indicators of the monitored air pollution condition. Moreover, additional KPIs are used to quantify the climate and vehicle traffic conditions aimed at enriching the analysis on pollutant concentration. The following KPIs have been defined.

(A) *Spatial-temporal pollutant concentration KPIs*. They report the pollutant concentration per sensor, at different time granularity levels in a user-specified date. The following three KPIs are defined. (A.1) Daily average pollutant concentration per sensor. (A.2) Average pollutant concentration per daily time slot (morning, afternoon, evening, night) per sensor. (A.3) Average hourly pollutant concentration per sensor.

(B) *Critical pollutant concentration KPI*. This KPI analyses how frequently the daily pollutant concentration per sensor is critical. Specifically, it reports the percentage of days within a given time period in which the pollutant concentration was critical according to a user-specified threshold.

(C) *Climate KPIs*. They report the hourly average values in a given date for various climate data as temperature, humidity, and pressure.

(D) *Vehicle traffic KPIs*. To characterize traffic under different perspectives, vehicles have been categorized based on the type of fuel (e.g., diesel, petrol, or methane) and the use of the vehicle (e.g., bus or private). KPIs report the total number of vehicle per day for each of the above categories, within a user-specified time period in a given urban area.

(E) *Statistical correlation analysis KPI*. Pollutant concentrations can be affected by meteorological conditions. To deepen the evaluation of the impact of climate conditions on air pollution, we analysed the statistical correlation between pollutant concentration and climate data. The established Pearson correlation coefficient [6] has been currently adopted for KPI computation over the sequence of measurements collected within a given time window. The Pearson correlation coefficient assumes values in the range [-1, +1]. Values higher than 0 represent positive correlations, while values lower than 0 represent negative correlations. The higher the absolute coefficient value the stronger the correlation.

APA includes four types of informative dashboards each one reporting one or more of the above KPIs. Dashboards have been designed to address the following targeted analyses: the analysis of (1) the *monitoring station characteristics*, (2) the *spatio-temporal pollutant concentration*, (3) the *impact of climate conditions on pollutant concentration,* and (4) the *impact of vehicle traffic on the pollutant concentration*. In the dashboards, we combined an interactive map to easily visualize the distribution of KPI values in geographical areas, with charts reporting more in detail the trend of KPI values over time. Dashboards are briefly described below.

(1) The *monitoring station characteristics* dashboard shows on a map the position of the pollution monitoring stations, and for each of them the types of measured pollutants. This dashboard provides useful information to understand the available hardware infrastructure, to analyze the statistical significance of the achieved results, and to support the planning of further extensions of the monitoring station network.

(2) The *spatial-temporal pollutant concentration* dashboards allow characterizing the pollution in urban areas over time. The temporal/spatial-based analysis allows understanding if some time periods/urban areas are more subject to air pollution than others. To support data analysis at different temporal granularity levels, three different dashboards have been defined based on KPIs on spatial-temporal pollutant concentration (KPIs (A)). An additional dashboard points out the number of days with pollutant concentration critical with respect to a user-specified threshold (KPI (B)).

(3) The *impact of climate condition on pollutant concentration* dashboards allow analyzing which weather conditions strongly affect the air quality in different urban areas and over time. A first dashboard includes the spatial-temporal analysis of the climate KPI trend jointly with pollutant concentration KPIs (i.e., KPIs (A) and (C)). A second dashboard reports the statistical correlation between climate conditions and pollutant concentration (i.e., KPI (E)).

(4) The *impact of vehicle traffic on pollutant concentration* dashboards allows analyzing how the circulating vehicles in urban areas impact on the concentration of the pollutants. To this end the vehicle traffic KPIs jointly with spatial-temporal pollutant concentration KPIs are computed and their trends (over time) are shown in easy readable charts. (i.e., KPIs (A) and (D)).

IV. DEVELOPMENT AND RESULTS

This section presents the current implementation of the APA system and its validation on real open data collected in Milan, a major city in the north of Italy. Some examples of interesting analysis scenarios are also discussed.

A. The APA development

The APA architecture is based on a multi-tier structure for supporting flexible, easily reusable and customizable applications. APA has been developed using open source technologies. Since all data are stored in the data warehouse based on the (fixed) schema reported in Figure 1, the technological solution adopted in APA for data storage exploits a relational DBMS. Specifically, MySQL is the RDBMS currently selected for empowering APA analytics. To visualize the results of the analysis, a secure responsive web-based application has been developed. The application supports various devices with a different screen size like mobile phones, tablets, and desktop computers. The application is based on HTML 5 and CSS 3 and it has been developed using the PHP programming language. To provide a more interactive navigation for the end user, the KPI visualization on geographical maps is based on a multiple-layer representation, where more geographical layers are overlapped (as the layer representing different geographical areas). Google Fusion Tables are used because of their potential in mapping KML files for creating layers on Google Maps.

In this study the assessment of the APA system was performed by deploying it on a dual-core 2.50 GHz Intel(R) Xeon(R) workstation with 4 GBs of RAM, running Ubuntu Linux 12.04 LTS.

B. Datasets

This section presents the three data sources considered in this study. Data from these sources have been collected for year 2013, and then integrated in APA as discussed in Section III.B.

The *ARPA Lombardia Pollutant dataset*[2] includes the concentration values for a set of pollutants (e.g, PM10, PM2.5, CO, O3) gathered through some monitoring stations located in the Lombardia Region. Each station is equipped with a set of sensors, each one measuring a given pollutant. We focused our analysis on the monitoring stations located in the city of Milan. The provided data consists of a set of hourly or daily readings, depending on the type of pollutant. Each reading is characterized by the monitoring station identifier, the sensor identifier, the name of the measured pollutant, the measured concentration of the pollutant, and the date and hour of the reading.

The *Weather Underground dataset*[3] includes the meteorological measurements collected through all Personal Weather Station (PWSs) registers by users. Since our scenario analysis is related to the city of Milan, we selected three PWSs near to the considered urban environment. For each selected PWS, meteorological measurements considered in APA (e.g., humidity, and atmospheric pressure) have been downloaded for the considered time period. Each record includes the identifier of the PWS, the timestamp of the measurement and for each weather measure the corresponding measured meteorological value. As discussed in Section III.B, we properly aggregated the weather data collected through different PWSs to obtain a representative value at the hourly granularity associated to each sensor of the pollutant monitoring stations.

The *Municipality of Milan's Traffic dataset*[4] provides an aggregated information about the number of vehicles entering in the central area of Milan per hour. The dataset contains the total number of entering vehicles per hour and the number of vehicles separately for each vehicle category. Vehicles are classified according to the fuel type (e.g., petrol, diesel, gasoline) and to the vehicle type (e.g., private, bus, goods). As discussed in Section II.B, traffic data has been associated to sensors monitoring pollutant concentrations in the central area of Milan.

---

[2] http://www2.arpalombardia.it/
[3] http://www.wunderground.com/
[4] http://dati.comune.milano.it/

C. *Analysis scenario*

Here we discuss four representative examples of interesting analyses that can be performed using APA. These analyses have been carried out on data related to the city of Milan, selected as a reference case.

Example (A): Analysis of daily pollutant concentration. The "Daily pollutant concentration" dashboard allows the user to select a date of interest and a pollutant to analyze. For the selected pollutant, the dashboard visualizes the following information. (i) The average concentration in the selected date in Milan areas, through colored markers placed on a map (one marker for each monitoring sensor). Markers are colored based on some predefined thresholds to visually represent a "good", "non-critical", "warning", "alert", or "alarm" pollutant concentration. (ii) The concentration trend over a ten days period including the selected date and the nine preceding dates. This analysis can support in identifying critical daily pollutant concentrations, but also in evaluating increasing/decreasing trends in pollutant concentration over a time period. As an example, Figure 2 reports the average daily PM10 concentration for April 11, 2013. In the upper part of Figure 2, the yellow markers on the map highlight a potential critical PM10 concentration on April 11, 2013 in three different areas of Milan. Moreover, the chart in the bottom part of Figure 2 shows an increasing PM10 concentration for all the three areas in dates from April 9 to 11, 2013, with the highest value on April 11 (bars on the right-end side of the chart). In the considered time period, a great event took place in Milan that may be one of the factors contributing to the increment of pollutant concentration, because a larger number of vehicles may have accessed the Milan areas.

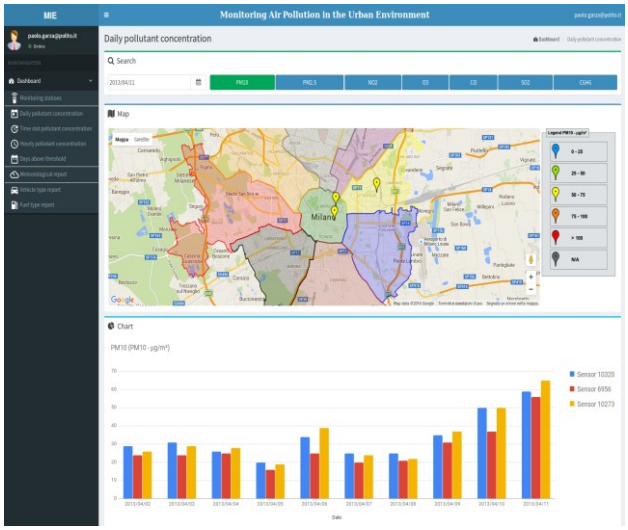

Figure 2 - Daily pollutant concentration dashboard

Example (B): Days above threshold. Using APA as discussed in Example (A), the user can analyse daily pollutant concentrations and discover critical days for pollutant concentration. Once a critical date, or a critical time period, has been pointed out, the user can deepen the analysis through the "Days above threshold" dashboard. This dashboard allows analyzing the percentage of days for which the pollutant concentration was above a user-specified reference threshold. This analysis can reveal the persistence of high pollutant concentrations over a long time period, which is critical for the wellness of citizens. As an example, we analysed the PM10 concentration from April 9 to 14, 2013, by setting 50 $\mu g/m^3$ as a threshold. The results show that in the analyzed time period in two out of the three sensors the percentage of days above threshold is higher than 33% and in the third sensor is 16%.

Example (C): Impact of vehicle traffic on pollutant concentration. The number and type of circulating vehicles can significantly affect the air quality. APA includes the "Impact of mobility on air pollution" dashboard to analyse the correlation between the number of vehicles in a given area and the pollutant concentration in the area, as well as to point out the most impacting types of vehicles on pollutant concentration. In the dashboard, first the user selects a time period of interest and a pollutant to analyze. Then, a chart in the upper part of the dashboard reports the number of vehicles entering in the central area of Milan each day in the time period. Both the total number of vehicles, and the number of vehicles per type, are reported. In the bottom part, for the same time period, a chart reports the daily concentration for the selected pollutant. As an example, Figure 3 reports the status of the dashboard for the PM10 pollutant in the time period from June 1 to 15, 2013. These results show that, for the analyzed area, (i) the majority of the traffic is related to private vehicles (approximately 80%) and (ii) the PM10 concentration has a trend similar to that of the number of private vehicles. For reducing the number of private vehicles, staff of the public administration can plan campaigns in favor of alternative (green) transportation vehicles (e.g., bicycles, electric cars, underground) for private users.

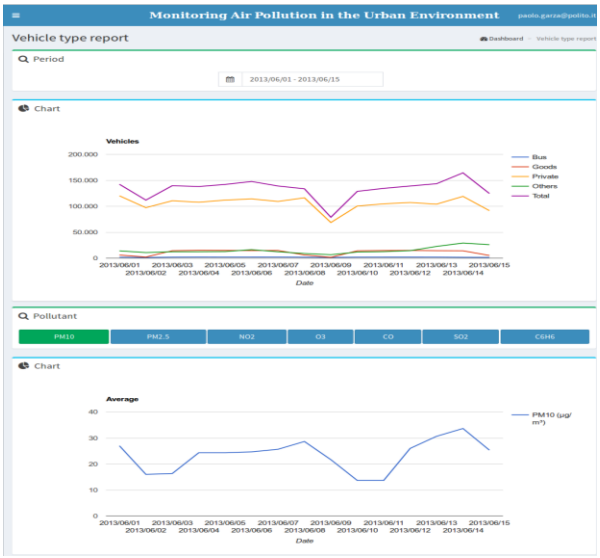

Figure 3 - Vehicle type dashboard

Example (D): Correlation between climate conditions and pollutant concentrations. This dashboard allows analyzing the statistical correlation between meteorological conditions and pollutant concentrations over a time period. Given a pollutant and a meteorological feature (e.g., temperature), the dashboard plots their statistical correlation (based on KPI (E)) per month within a selected time period. For example, we analyzed the correlation between temperature and ozone from January, 2013 to December, 2013. The chart (not reported here for sake of space) showed a positive correlation over all time period (correlation value always greater than 0) and a significant positive correlation from June to September (correlation value greater than 0.5) when the temperature is usually higher (i.e., in the summer season). This result is consistent with previous studies discussing the impact of temperature on ozone concentration [7].

V. CONCLUSIONS AND FUTURE WORKS

This paper presented APA, a business intelligence engine to analyze air pollution from different perspectives (e.g., pollutant concentration, meteorological data, traffic data) for the purpose to make citizens aware of air quality and to understand how people activities impact on air quality. As a case study, APA has been evaluated on open data on air pollution in the urban area of a major Italian city. There is still room for improvements for our system. For example, APA may be enriched with advanced data mining algorithms: (i) to analyze the correlations hidden in the air pollution-related data at different abstraction levels [8] and (ii) to forecast fine and coarse grained air pollution data [9].

VI. ACKNOWLEDGMENT

The research leading to these results has received funding from the Italian Ministry of Research (MIUR) under the "Cluster Tecnologie Smart Communities Progetto MIE – Mobilità Intelligente Ecosostenibile".

REFERENCES

[1] Y. Zheng, F. Liu, and H. Hsieh, "U-air: when urban air quality inference meets big data," in the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013, pp. 1436–1444, 2013.

[2] S. Devarakonda, P. Sevusu, H. Liu, R. Liu, L. Iftode, and B. Nath, "Real-time air quality monitoring through mobile sensing in metropolitan areas," in Proceedings of the 2Nd ACM SIGKDD International Workshop on Urban Computing, pp. 15:1–15:8, 2013.

[3] Y. Xiao and C. Ji, "Management of air quality monitor data with data warehouse and GIS," in CSIE 2009, 2009 WRI World Congress on Computer Science and Information Engineering, pp. 148–152, 2009.

[4] Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, "Forecasting fine-grained air quality based on big data," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2267–2276, 2015.

[5] M. Golfarelli, D. Maio, and S. Rizzi, "The dimensional fact model: A conceptual model for data warehouses," in International Journal of Cooperative Information Systems, vol. 7, pp. 215–247, 1998.

[6] J. Han, "Data Mining: Concepts and Techniques," Morgan Kaufmann Publishers Inc., 2005.

[7] E. Stathopoulou, G. Mihalakakou, M. Santamouris, and H. S. Bagiorgas, "On the impact of temperature on tropospheric ozone concentration levels in urban environments," in Journal of Earth System Science, vol. 117, issue 3, pp. 227-236, 2008.

[8] D. Antonelli, E. Baralis, G. Bruno, L. Cagliero, T. Cerquitelli, S. Chiusano, P. Garza, and N. A. Mahoto, "Meta: Characterization of medical treatments at different abstraction levels," ACM TIST, vol. 6, no. 4, p. 57, 2015.

[9] E. Baralis, T. Cerquitelli, S. Chiusano, A. Giordano, A. Mezzani, D. Susta, and X. Xiao, "Predicting cardiopulmonary response to incremental exercise test," in 28th IEEE International Symposium on Computer-Based Medical Systems, CBMS 2015, pp. 135–140, 2015.