

This is a pre print version of the following article:

Visual analytics for spatio-temporal air quality data / Bachechi, Chiara; Desimoni, Federico; Po, Laura. - (2020). (Intervento presentato al convegno 24 International Conference Information Visualisation (IV2020) tenutosi a Victoria University, Australia & Technische Universität Wien, Austria nel 7-11 September 2020).

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

27/04/2024 15:09

Visual analytics for spatio-temporal air quality data

Chiara Bachechi*, Federico Desimoni*, Laura Po* and David Martínez Casas†

*‘Enzo Ferrari’ Engineering Department, University of Modena and Reggio Emilia, Italy

Email: name.surname@unimore.it

†CiTIUS, Universidade de Santiago de Compostela, Spain.

Email: david.martinez.casas@usc.es

Abstract—Air pollution is the second biggest environmental concern for Europeans after climate change and the major risk to public health. It is imperative to monitor the spatio-temporal patterns of urban air pollution. The TRAFair air quality dashboard is an effective web application to empower decision-makers to be aware of the urban air quality conditions to define new policies and keep monitoring their effects. The architecture copes with the multidimensionality of data and the real-time visualization challenge of big data streams coming from a network of low-cost sensors. Moreover, it handles the visualization and management of predictive air quality maps series that is produced by an air pollution dispersion model. Air quality data are not only visualized at a limited set of locations at different times but in the continuous space-time domain, thanks to interpolated maps that estimate the pollution at un-sampled locations.

Index Terms—air quality, temporal data, spatial data, visualization, interpolation maps, prediction

I. INTRODUCTION

Nothing is more vital to life than breathing. People spend a lot of time deciding what to eat or drink to be healthy, they should also consider the influence on their lifestyle of the quality of the air they breathe. Over 90% of the world’s population lives in places where air pollution is above the World Health Organization (WHO) air quality guidelines (AQGs) [1]. Pollution in the air is invisible but it kills 7 million people a year, far more than HIV, tuberculosis, and malaria combined. In Europe, 500,000 deaths per year are caused by air pollution. To face this issue public administrations and citizens need systems able to let them visualize the entity of the problem to change their behavior and adopt effective policies. The TRAFair project¹ aims to monitor urban air quality in real-time and to predict air pollutant concentrations for the next two days. Air pollution depends on the concentration of certain substances recognized as toxic. If the concentration of these substances is higher than the legal limit for a while to cause harm or undesirable effects the air is considered polluted. Thus, the concentration of pollutants in the air is the value to measure and compare with European standards. Within TRAFair, a network of low-cost air quality sensors has been deployed in 6 cities to monitor air conditions. These sensors can measure the concentration of four pollutants: CO , NO_2 , NO , and O_3 . The collected information is spatio-temporal data and thus require visualization techniques to represent both dimensions. This paper describes the TRAFair Air Quality

dashboard that has been implemented to share and visualize air quality data for the city of Modena. The dashboard provides a set of visual analytics - dynamic and/or interactive graphics to display and communicate information about real-time, statistics, and trends of air quality conditions in a city. The main features of the dashboard are: (1) it can manage a large amount of data coming in real-time from different sensors and characterized by both spatial and temporal dimensions, (2) it can handle complex data such as series of maps obtained from models or interpolation process, and (3) it is scalable and easily adaptable to different cities. The rest of the paper is structured as follows. In Section II, a literature review on air quality data visualization is provided. Moreover, in Section III, some background knowledge to better understand the air pollution monitoring system is given. In Section IV, the system architecture that enables the visualization is described. Then, in Section V, dashboard properties, charts, and visualizations are presented. Finally, the results displayed in the dashboard are commented and future works are described in section VI.

II. RELATED WORK

In [2], an analysis of city dashboards is provided considering mainly geo-visual analytics to assess the creation of design principles for communicating real-time data focusing on time-series real-time data. The paper discusses the main real-time related challenges: change blindness, and communication of spatio-temporal variability. In the realization of the TRAFair dashboard, the presented principles have been taken into account. In [3], a classification of different types of dashboards used for smart cities is provided. They describe geospatial dashboards as a web-based interactive interface supported by a platform combining: mapping, spatial analysis, and visualization. Three different types of dashboards are distinguished: Operational dashboards that provide descriptive measurements of smart cities using indicators, Analytical dashboards that are a diagnostic method for smart cities based on data inferred from geospatial data using spatial analytics, and Strategical dashboards that are predictive dashboards used by smart cities to estimate and visualize possible future outcomes. We implement our dashboard considering the description given in [3] to create a unique platform that can be an operational, analytical, and strategical dashboard at the same time.

¹<https://trafair.eu>

III. BACKGROUND

In the city of Modena, within the TRAFair project 13 low-cost air quality sensors have been installed [4]: 12 Decentlab Air Cubes and one Libelium Smart Environment PRO. These sensors are connected to the municipality LoRaWAN network and send concentrations of 4 gases every 2 minutes.

The sensors are moved around the city in different locations during their life cycle. Firstly, they need to be calibrated. The calibration requires a period of time in which the sensors are collocated near an air quality legal station. In Modena, there are two legal stations managed by ARPAE the environmental agency of the Emilia-Romagna region. Every 6 months, the calibration process is replicated to ensure the quality of the observations produced by the sensors. Once calibrated, each sensor is able to provide a value of concentration for every pollutant.

In order to present the air pollutant concentrations to diverse audiences, we decide to share numeric data and to display them using two color scales: the one used by the ARPAE, and the one proposed by the European Environmental Agency (EEA). Public administrators need to know the position and status of the installed sensors, to monitor real-time air quality conditions, to visualize statistics concerning the air quality conditions in the urban area, and to quickly identify if there are hot-spots, i.e. areas with a high concentration of pollutants. Each air quality sensor provides the four pollutant concentrations in the location where it is placed. By spatially interpolating these values, it is possible to estimate the pollutant concentration in the whole urban area. An R script was implemented to produce GeoTIFF files interpolating the pointwise concentrations using Inverse Distance Weighted (IDW). As described in [5], IDW is a deterministic (non-geostatistical) estimation method where values at unmeasured points are determined by a linear combination of values at nearby measured points. Value at location x^* is evaluated as:

$$x^* = \frac{w_1 X_1 + w_2 X_2 + \dots + w_n X_n}{w_1 + w_2 + \dots + w_n}$$

where x^* is the value to predict and w_i is the weight of the sampled point X_i . Weights are evaluated as the inverse of the distance between the location to predict and the sampled data point i to the power of p :

$$w_i = \frac{1}{d_{i-x^*}^p}$$

The rate at which the weights decrease is dependent on the value of p . As p increases, the weights for distant points decrease rapidly. If the value of p is very high, only the immediate surrounding points will influence the prediction. In our use case, we employ $p=2$. Other interpolation techniques such as IDW with p equal to 1, Nearest neighborhood, Ordinary Kriging, and Thin plate spline, have been considered and tested. Both IDW and Nearest Neighbourhood provided good interpolated maps. We chose to use IDW and to produce a new map every time new calibrated values are available i.e. every 10 minutes.

An urban air pollution dispersion model to predict the quality of urban air based on weather forecast and traffic flows has been set up in each city involved in TRAFair [4]. The air pollution dispersion model used as input weather forecast data and atmospheric emissions data and produce as output a dispersion map for the urban areas that shows pollution concentration values on an urban grid of 2-4 meters for every hour of the following two days. The open-source simulation software Graz Lagrangian Model (GRAL) is run each day in Modena to predict the air quality conditions in the next 48 hours [6]. This model required to input the weather forecasts, wind directions, vehicle emissions, and the shape of the buildings in the urban area. Every day, the GRAL model generates 48 GeoTIFFs, one for each forecasted hour, representing the concentration of NOx in the urban area.

IV. ARCHITECTURE AND TECHNOLOGIES

Figure 1 depicts the architecture of the dashboard. The data are stored in the TRAFair DB. The information contained in the database is exposed through GeoServer over the Internet and used by the air quality dashboard.

A geospatial dashboard is not only a tool for displaying and visualize geospatial data, but it also supports data mining and decision-making. Since smart cities are ingesting more and more data and the urban sensor network is destined to increase, one of the necessary features of a geospatial dashboard is scalability. Scalability means the ability to add or remove sensor devices without affecting the system availability.

A. TRAFair database

The database collects the stream of data generated by the low-cost air quality sensors. At first, the database contains basic information about the sensors themselves. Then, every time the sensors are moved, new data related to its current position and status are inserted. At one time, a sensor is located in one place and can be in one of the following statuses: *running*, *calibration*, *broken*, and *offline*. For each observation captured by the sensor, we store the measurement voltages of each gas (in millivolt), the temperature, the humidity, and the battery voltage. Once the calibration step is complete, the data incoming from the sensors, are converted from millivolts to concentration (milligrams for cubic meter) and then stored in the database. We adopted a PostgreSQL database with the PostGIS and timescale extensions for the efficient management of the time series and spatial data. Alongside with the schema containing the core database, we created 3 new partitions that contain tables and views with aggregated data to create statistics and visualizations. In particular, we deployed the web app schema to store views used by the web dashboards.

B. GeoServer

GeoServer² [7] is an open-source server for sharing geospatial data. GeoServer allows us to publish and upload different types of data sources: a file or group of files (raster or

²<http://geoserver.org/>

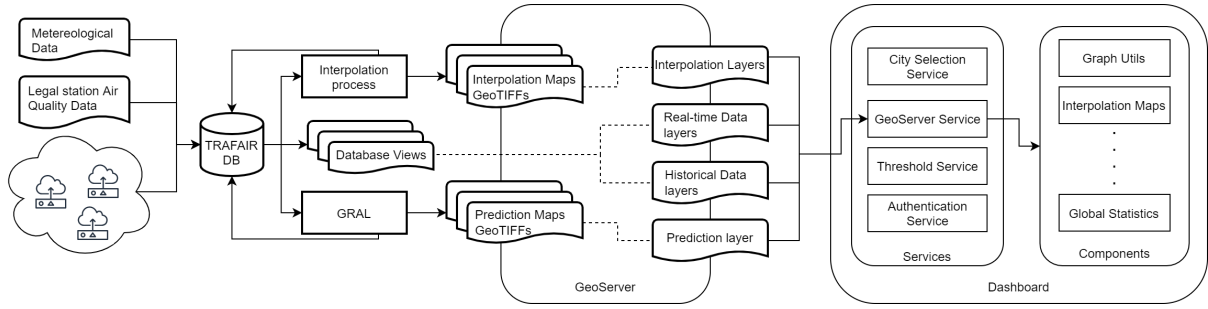


Fig. 1. Architecture of the Air Quality Dashboard.

vector data), a table in a database, a single raster file, or a directory. Each data source is connected to a store. Then, from each store, different layers can be created. Through layers, GeoServer enables the handling of temporal and spatial data using standard OGC protocols such as the WFS (Web Feature Service), the WCS (Web Coverage Service), and the WMS (Web Map Service). Through the WFS, it is possible to expose spatial and temporal data as GML, KML, JSON, CSV, XML, and other data formats; indeed, we use this protocol to expose the data that will be used for drawing the graphs that will be presented in Section IV-C. The WCS is used for exposing “coverages” which are objects covering a geographical area like a set of points, a regular grid of points, or a set of segmented curves. The WMS is used for delivering map images, it generates server-side maps and sends them to the clients as regular images, no fine-grained data is sent to the client; we use this protocol for exposing the interpolation maps described in Section V-C and the prediction maps generated by the air pollution dispersion model described in Section V-D. We deployed an instance of GeoServer to achieve a higher security level (by separating the database from the visualization layer) and to manage easily the publication of spatial and temporal data. Moreover, GeoServer has been equipped with the ImageMosaic plugin. The ImageMosaic data store allows the creation of a mosaic from a set of GeoTIFFs. This operation is used to assemble a set of geospatially rectified images, i.e we use it to group the GeoTIFFs prediction map of several time slots into a unique element. Finally, GeoServer provides an API system that simplifies the automation process. We exploited this API system to generate the layers required for exposing the database’ views automatically, similarly to what has been done for the release of air quality open data depicted in [8].

For handling sensor observations, status and positions, and statistical data about sensors or locations, we realized 9 views in the web app schema and produced 9 layers in GeoServer. These layers are grouped in 3 Real-time Data layers and 6 Historical Data layers as shown in Figure 1. Moreover, to manage the huge amount of data, some views have been materialized. The advantages result in an average response time that decrease from 10 to 0,7 seconds. The 4 interpolation map layers are built from the GeoTIFFs created by the interpolation process where each layer contains a GeoTIFF

file. The prediction map layer, instead, is created exploiting the ImageMosaic plugin to merge the 48 GeoTIFFs generated by GRAL into an ImageMosaic data store. The content of the Historical data layers and Prediction layer is refreshed daily while the Interpolation layers and real-time data layers are refreshed every 10 minutes.

C. Dashboard

The dashboard is a web application implemented using Angular 7 application design framework and Typescript language described in [9]. D3 [10] and Chartjs [11] are the JavaScript libraries used to show graphs. Open Layer instead is the library employed to create maps. Angular is a platform for building single-page client applications. Angular allows reusing of code thanks to its modular structure. Angular code is based on building blocks called NgModules that collect related codes into functional sets. These modules provide the compilation context for ‘components’. Components are composed of several screen elements. The functionalities that are not related to views are provided by ‘services’. Components can use all the available services and can also employ them to share data, information, and functionalities with other components. Each view of our application has a corresponding component, some of these components have a parent-child relationship. This means that the child component is inserted inside the parent component view. Properties in a child component can receive their values from their parent component using the ‘@input’ decorator. In our application some views contain graphs that change accordingly to some options selected by users, graphs are child components of these views.

We define several services to share information between components, these services support both the retrieval and the modification of the information. The main services we developed, as shown in Figure 1, are:

- Authentication service: to share and modify data (permissions, status) concerning the user that logged in.
- City selection service: to share and modify data about the selected city (our web application supports three cities)
- GeoServer service: to share functionalities related to GeoServer interrogations.
- Threshold service: to share and modify the selected color scale thresholds that determine the graphs and maps colors.

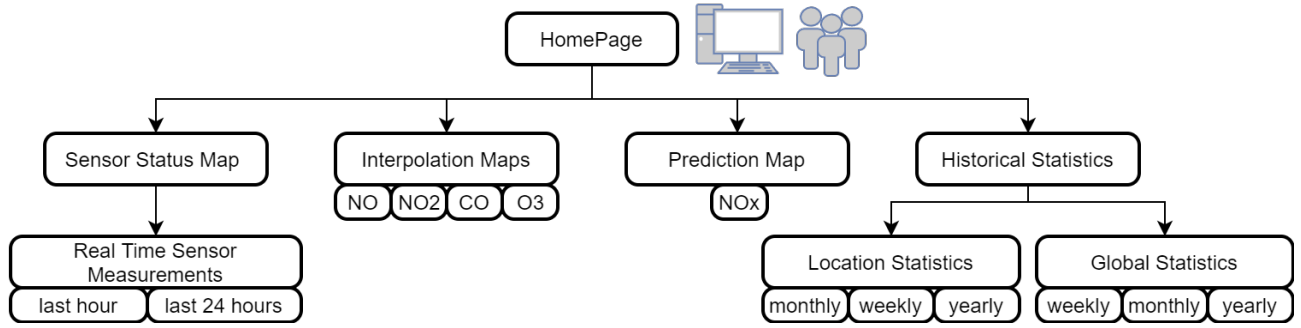


Fig. 2. Site Map of the TRAFair Air Quality Dashboard.

V. TRAFair AIR QUALITY DASHBOARD

The TRAFair Air Quality Dashboard³ is part of a suite of monitoring tools for public administrations developed within the TRAFair project. It enables decision-makers to monitor a large amount of quickly transitioning data.

The dashboard provides means for the analysis and the estimation of the diffusion of pollutants in the urban area, allowing users to track and compare over time and space, and in case of real-time data, the here-and-now, of air quality conditions in the urban area. The dashboard employs mainly graphs and maps to display time series and spatial data. The graphics are dynamic, i.e. they are updated as new data are released. Some of them are interactive through operations such as selecting, filtering, and querying data, zooming in/out, panning, and overlaying. The dashboard has been conceived for public administrations to help them in deciding for the future of the city, not all the visualized data are public. The source of city maps is Open Street Map (OSM)⁴, the data visualized are obtained querying GeoServer configured layers. For real-time data visualization, timers have been inserted inside the Typescript code of the web application to control the frequency of updates of views. When the timer associated with a view expires the GeoServer layer is queried for new data, the view is updated, and the timer is restarted. According to Figure 2, the site map is organized in 4 branches, in the following sub-sections, we are going to describe the view displayed in each branch.

A. Sensor status and real-time measurements

The dashboards facilitate the exploration of multiple sensors' real-time measurements simultaneously. Firstly, in the 'sensor map' view, the position of each sensor and its status is displayed (see Figure 3-1). Each status is associated with a different color; the sensor map shows a spot for each sensor colored accordingly to its status. Some sensors are placed in the same location: this happens when they are in calibration status since the need to be placed near a legal station. To enable the view of sensors in the same locations, the Open Layer library has been modified to manage clusters of sensors.

The map, initially, shows a bigger spot displaying the number of sensors located in the specific position, and then by clicking on the cluster spot, a circle of spots appears allowing the user to see the status of each sensor (see Figure 3-2).

The user can interact with the map by clicking on a spot; this action opens a new view where the last 24 hours measurements of the 4 pollutants of the selected sensor are displayed (see Figure 3-3).

To display real-time measurements, a linear representation of time is used, this solution demands greater display space to produce a comprehensible visual presentation. For this reason, by clicking on a graph, referring to a specific pollutant, the graph is zoomed and the color of the background change accordingly to the used color scale described by the legend (see Figure 3-4). The color scale can be chosen between two options (ARPAE and EEA), the available color scales are described in the setting-up page. Once the user chooses a color scale, the threshold service will provide this information to all the views in the application. At the top of the page, there is a button that allows us to display only the last hour measurements (see Figure 3-5). These views showing real-time measurements are considering only the time dimension of sensor observations. To take into account also the spatial dimension and allow the user to easily compare the measurements of different sensors in different locations an additional view has been added. As shown in Figure 4, this view contains a small map that shows the position of each sensor; where all sensors are colored differently. Then, for each pollutant, the time series of sensor measurements are displayed in the same chart. The color of the curve associated with a sensor is the same color of the sensor spot in the map. The user can interact with these graphs clicking on the label of a sensor to remove its measurements and compare a defined group of sensors. This view combines spatial and temporal dimension in a single visualization. Every 10 minutes, the graphs are automatically refreshed.

B. Historical Statistics

Historical data are collected in the TRAFair database and uploaded in GeoServer forming a longitudinal archive of data that can be examined over different time frames, such as a week, month, and year. Archival data are used to

³The dashboard is available online at <https://trafair.eu/airquality>. Credential of the test account: user=test, pwd=test

⁴<https://www.openstreetmap.org/>

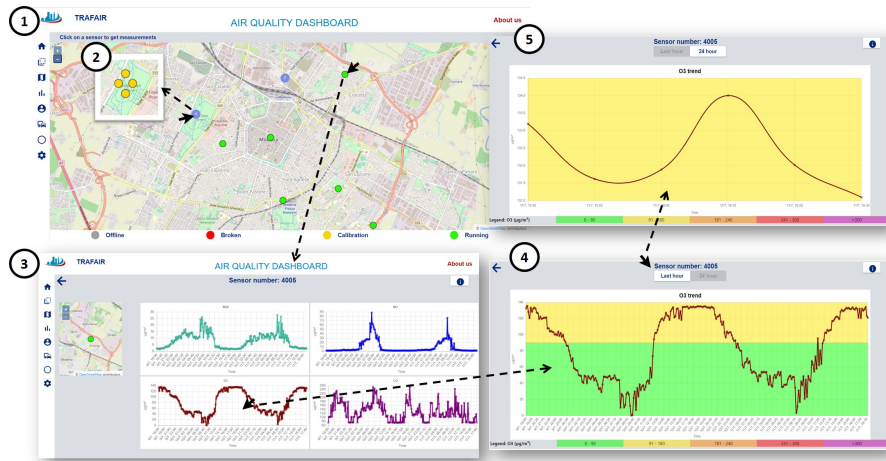


Fig. 3. All views accessible from the sensor map view.

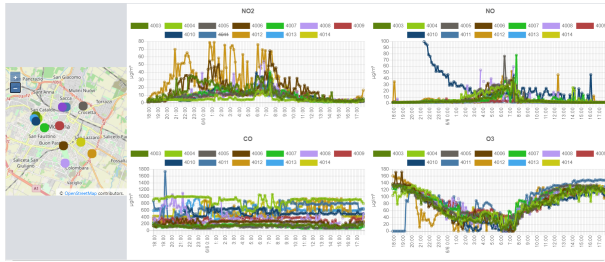


Fig. 4. All views accessible from the sensor map view.

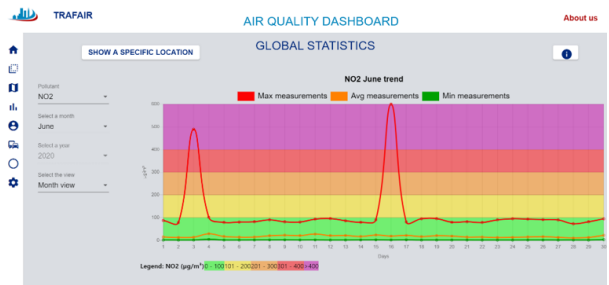


Fig. 5. Month view for NO_2 , June 2020. The colors of the background is based on ARPAE scale.

establish trends and providing contextual information for a better understanding of current data. The interactive display allows the viewer to select the location, the month, and the year changing the display of statistical data to reflect the selected time aggregation. Two different types of statistics are provided: Global statistics, and location statistics. Global statistics are evaluated considering all sensors available in the corresponding period of time, they can provide information about the general air condition in the city. Location statistics are provided for each location of the city where sensors have been placed. Sensors are moved around the city, so it can happen that several sensors are in the same location. Since the air pollutant concentration does not depend on the physical sensor, the statistics are evaluated considering the observations registered by all the sensors that have been placed in that specific location during the observed period. The user interacts with a map showing all the available locations and by clicking on them he/she can visualize the view of the selected location statistics. For both global and location statistics, three different graphs are provided: the month view, the year view, and the week view. The month view is available for each month, it requires to select the month, the year, and the pollutant. The generated graph shows for every day of the selected month the maximum, minimum, and average time series of values of pollutant concentration measured in the days of the month. The



Fig. 6. Two year views for CO in 2020: the bottom uses ARPAE scale and the top EEA scale. The figure shows also the setting view where scales can be modified.

averages, minimums, and maximum curves can be removed by clicking on 'Max measurements', 'Avg measurements', and 'Min measurements'. The background colors indicate the level of concentration of the pollutant accordingly to the selected color scale. In Figure 5, the global month view of NO_2 for June 2020 is displayed. The year view is available for each year, the user can select the year and the pollutant, and then the visualized graph shows for every month of the

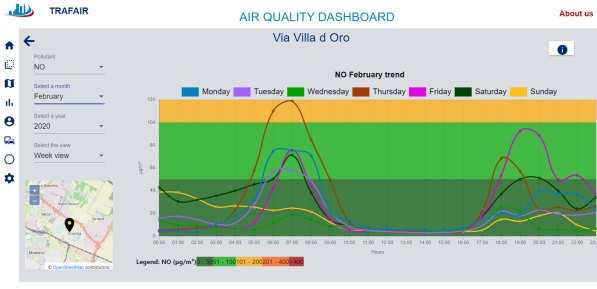


Fig. 7. Average weekdays view for O_3 in July 2020 in Via Villa d'Oro. The colors of the background is based on EEA scale.

selected year the maximum, minimum, and average value of the selected pollutant concentration measured in that month. The maximum, minimum, and average curves can be removed to ensure a better visualization, and the background is colored accordingly to the selected color scale. As an example, in Figure 6, the graphs for CO in 2020 is showed with the two different scales. It can be observed that the same values are located in different color classes. The week view is available for each month, once the month, the year, and the pollutant have been selected the graph is shown. The curves of the average concentration of the mean weekdays are displayed. For each weekday, all the available days for the same weekday in the selected month are averaged to get the mean hourly concentration trend. The curves of different weekdays are displayed together in the same graph to allow comparison but can be removed clicking on the weekday name. Also in this graph, the background is colored accordingly to the selected scale. In Figure 7, a specific location week view showing average weekdays trend of O_3 concentration for February 2020 is displayed. The view also displays a map that shows the position of the selected location.

C. Interpolation maps

In the interpolation maps view of the dashboard, GeoServer layers are queried to obtain the semi-real-time interpolation maps. The view, displayed in Figure 8, allows to select the desired pollutant and shows the corresponding colored interpolation map overlapped on the OSM city map of the area. The style of the map can be changed accordingly to the selected color scale. The time interval the map refers to is indicated in the left bottom of the view and the maps are updated automatically every 10 minutes.

D. Prediction map

As described in Section III, the GRAL dispersion model generates 48 images, for each hour of the forecast. These images are exposed through a GeoServer layer that is queried by the prediction map view allowing the user to use the scrollbar on the bottom to move through time and see the evolution of the air conditions during the two forecasted days as displayed in Figure 9.

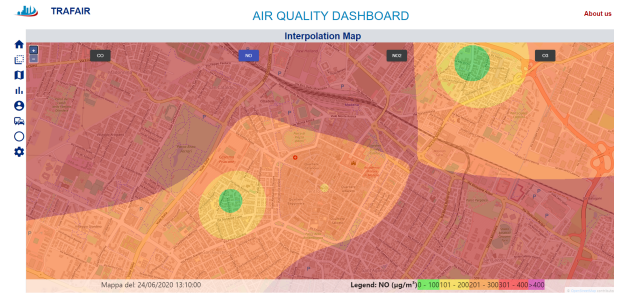


Fig. 8. Interpolation map of NO for the city of Modena. The colors are based on EEA scale.

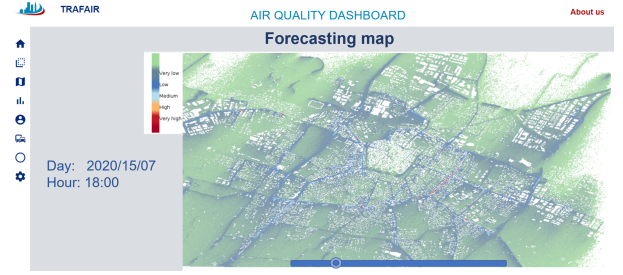


Fig. 9. Prediction view that displays the air quality forecast for 2020/07/15 at 18:00.

VI. CONCLUSION

The paper presented the TRAFair Air Quality dashboard that has been realized within the TRAFair project to display real-time, forecasted and statistical air quality observations through graphs, timelines, and maps. It allows decision-makers to monitor air conditions in the city of Modena and analyze trends regarding pollutant concentration. The architecture of the web application enables scalability allowing us to insert new sensors, to move sensors around the city in new positions, and collect statistics regarding more months and years. It also permits replicability in different cities. Through the use of GeoServer data can be queried as coming from an API with a REST request. This allows us to manage a big amount of data structuring them in different layers and enabling different services on the same storage to obtain images, maps, or numerical values. The response time is reduced by the use of materialized views and automatized processes to update them.

In the near future, the dashboard becomes a multi-city dashboard that will be used also in Santiago de Compostela and Zaragoza. We intend to integrate into the dashboard a combined visualization of air pollution dispersion model results and measured data to provide feedback on the effectiveness of the predictions to refine the model. Also, the impact on air quality of a new hypothesis of circulation (such as changes in the urban fleet of vehicles, an increase of low emission vehicles) will be shown in the dashboard.

ACKNOWLEDGMENT

This research has been supported by the TRAFair project 2017-EU-IA-0167), co-financed by the Connecting Europe Fa-

cility of the European Union. The contents of this publication are the sole responsibility of its authors and do not necessarily reflect the opinion of the European Union. The authors would like to thank all the partners that contribute to the collection and management of sensor data: the City of Modena and Lepida S.c.p.A. A special thanks go to the LARMA team for their invaluable work on the installation maintenance and calibration of air quality sensors. We would like to express our gratitude also to Jose R. R. Viqueira for his feedback on the paper and to Raquel Trillo Lado and Yolanda Castillo for their help in the creation of interpolated maps.

REFERENCES

- [1] W. H. O. (WHO), “Healthy environments for healthier populations: Why do they matter, and what can we do?” WHO reference number: WHO/CED/PHE/DO/19.01, Licence: CC BYNC-SA 3.0 IGO., 2019. [Online]. Available: <https://www.who.int/phe/publications/healthy-environments/en/>
- [2] S. Stehle and R. Kitchin, “Real-time and archival data visualisation techniques in city dashboards,” *Int. J. Geogr. Inf. Sci.*, vol. 34, no. 2, pp. 344–366, 2020. [Online]. Available: <https://doi.org/10.1080/13658816.2019.1594823>
- [3] C. Jing, M. Du, S. Li, and S. Liu, “Geospatial dashboards for monitoring smart city performance,” *Sustainability*, vol. 11, p. 5648, 10 2019.
- [4] L. Po, F. Rollo, J. R. R. Viqueira, R. T. Lado, A. Bigi, J. C. López, M. Paolucci, and P. Nesi, “Trafair: Understanding traffic flow to improve air quality,” in *2019 IEEE International Smart Cities Conference, ISC2 2019, Casablanca, Morocco, October 14-17, 2019*, 2019, pp. 36–43.
- [5] J. Li and A. D. Heap, “Spatial interpolation methods applied in the environmental sciences: A review,” *Environmental Modelling & Software*, vol. 53, pp. 173 – 189, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364815213003113>
- [6] A. Bigi, G. Veratti, S. Fabbi, L. Po, and G. Ghermandi, “Forecast of the impact by local emissions at an urban micro scale by the combination of lagrangian modelling and low cost sensing technology: The trafair project,” in *19th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Harmo 2019*, 2019.
- [7] S. Iacovella, *GeoServer Cookbook*. Packt Publishing, 2014.
- [8] J. Noguera-Iso, H. Ochoa-Ortiz, M. A. Janez, J. R. R. Viqueira, L. Po, and R. Trillo-Lado, “Automatic publication of open data from ogc services: the use case of trafair project,” 2020, manuscript submitted for publication in The Twelfth International Conference on Advanced Geographic Information Systems, Applications, and Services GEOProcessing 2020.
- [9] N. Jain, A. Bhansali, and D. Mehta, “Angularjs: A modern mvc framework in javascript,” *Journal of Global Research in Computer Science*, vol. 5, no. 12, pp. 17–23, 2014.
- [10] M. Bostock, V. Ogievetsky, and J. Heer, “D3 data-driven documents,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, p. 2301–2309, Dec. 2011. [Online]. Available: <https://doi.org/10.1109/TVCG.2011.185>
- [11] “Chart.js — open source html5 charts for your website,” <http://www.chartjs.org/#Dataviz>, 2017, accessed: 2020-07-15.