A Graphical Tool for Eliciting Knowledge of Air Pollution Sources

Duarte Vital*, Pedro Mariano*†‡, Susana Marta Almeida‡, Pedro Santana*†
* *Instituto Universitário de Lisboa (ISCTE-IUL)*, Lisboa, Portugal
Emails: duarte_vital@iscte-iul.pt, plsmo@iscte-iul.pt, pedro.santana@iscte-iul.pt
† *Information Sciences and Technologies and Architecture Research Center (ISTAR-IUL)*, Lisboa, Portugal
‡ *Centro de Ciências e Tecnologias Nucleares*, *Instituto Superior Técnico*, Lisboa, Portugal
Email: smarta@ctn.tecnico.ulisboa.pt

***Abstract*—With tens of thousands of air quality monitoring stations installed in the world, this source of information has become the standard in air quality measuring. Air pollution becoming a growing concern for decades now, the need for an easy way to visualise pollution data arose. Extensive maps have been created to represent air pollution using data collected from monitoring stations, as well as other sources of information, such as traffic density or weather forecasts. This paper introduces a complementary source: direct environmental expert knowledge. By using the developed tool, the goal is to allow experts to express their knowledge about air pollution emission and diffusion as a function of the presence of key topological elements in a map, such as buildings or roads. The results of the usability tests performed with a sample of 30 participants are promising. Participants provided useful feedback regarding key application features to be implemented in future iterations.**

***Index Terms*—air quality mapping, graphical educational tool, environmental awareness, environmental communication**

I. Introduction

Traditionally, air pollution monitoring has relied on complex and costly monitoring stations. As such, they are located in major cities. With the advent of low-cost sensors (LCS), there are now several networks of these sensors that provide more coverage and, thus, more information [1]–[3]. Although they are less accurate and reliable, they can still be used to complement monitoring stations [4]. Despite this deployment effort, there are still large regions of the globe, including cities, that are not monitored.

Various currently available pollution mapping solutions [5]–[7], while also using other sources of information, are often dependent on monitoring stations as data providers. This is the main issue with current solutions, given that there are not enough monitoring stations to properly represent the whole world's pollution distribution with a street-level accuracy. In areas where these data are not available, an accurate pollution map is difficult to build, as estimates based on other sources must be relied upon (for a review of different pollution modelling approaches the reader can refer to [8]). In fact, pollution maps generated from estimation models, rather than from sensory observations, are often incomplete. A model that does not perform land-use regression may miss out heavy polluters, such as coal factories.

This paper presents an alternative approach for pollution mapping when in the absence of sensory data. Concretely, this paper presents a graphical tool that performs knowledge elicitation from environmental experts and applies the gathered knowledge to produce dynamic pollution maps for a given geographical location.

In the presented tool, expert knowledge is represented as associations between city elements (e.g., buildings, roads) in a given city map and pollution emission and dispersion patterns. These associations are created by the expert via the graphical tool. Then, these associations are used by the tool to generate dense and dynamic pollution maps. For instance, given the previous example, an expert can easily identify a coal factory as a source of pollution. In particular, the expert can select elements and assign them pollution profiles. These can depend on hour, weekday, or any parameter that the expert desires (e.g., main wind direction). Pollution is represented by a dynamical heat map.

The interaction metaphors and the look-and-feel of the tool's interface were influenced by existing map-based applications. As an example, there is the interface design of Google Maps and the way information is presented in each layer (standard, satellite imagery). Another example is Open Street Map (OSM) which is a popular library to display geographical data. The goal is to reduce the steepness of the learning curve. Functionality and data availability were key aspects in selecting a particular map library.

The main focus of this paper is on the ability for environmental experts to express their knowledge using the devised tool. To assess the usability of the developed tool and the viability of the approach, a set of tests with thirty participants was carried out. These participants were asked to perform several tasks with the tool and to answer a usability questionnaire. The obtained results show the usability of the tool and the merits of the proposed expert-based pollution mapping approach.

The research reported here is part of the ExpoLIS project. Our goal in the ExpoLIS project is to deploy a network of mobile LCS to monitor air quality and to develop a set of software tools [9]. The goals of these software tools include

This work was developed in the scope of the project ExpoLIS (LISBOA-01-0145-FEDER-032088) funded by FEDER and by national funds, through FCT – Foundation for Science and Technology, IP. C2TN authors also acknowledge the support of FCT for funding the strategic project UIDB/04349/2020.

TABLE I
MEANING, RANGE OF VALUES AND COLOUR OF AIR QUALITY INDEX LEVELS.

Health Concern Level	Value Range	Colour
Good	0-50	
Moderate	51-100	
Unhealthy for sensitive groups	101-150	
Unhealthy	151-200	
Very unhealthy	201-300	
Hazardous	301-500	

improving the work flow of an environmental expert and to increase citizens' awareness of air pollution. The tool presented in this paper is an example of the first goal. As for the second goal, we have developed a mobile app for easy access to the generated air quality data [9] and a 3D virtual environment for immersive and engaging air quality data visualisation [10].

The article is organised as follows. Section II surveys related work. Then, the developed tool is presented in Section III. In Section IV, the testing phase is detailed and the obtained results are presented and analysed. A set of conclusions and future work direction are presented in Section V.

II. RELATED WORK

Disparate air pollution measuring units, with different ratings, criteria and representations, are employed by countries and existing mapping solutions. One of the most often used is the American version of the Air Quality Index (AQI) [11], particularly in scientific studies [12]–[15]. The AQI converts pollutant concentration values into a colour-coded scale, in which higher values indicate increased health risk, as depicted in Table I.

Two main types of pollution mapping solutions are most prominent. The first type generates maps that display monitoring stations' locations and respective gathered air pollution data. The World Air Quality Index [16] is a prime example of a project that relies on this type of map. It collects data from more than 30,000 monitoring stations in 2,000 major cities containing individual pollutants concentrations and, through these, an AQI value is obtained. By inspecting the generated maps, it becomes clear that not enough monitoring stations are currently installed in order to represent world-wide air quality at a block-level precision, which is mostly due to their high acquisition and maintenance costs.

The second type of pollution mapping solutions produce estimated pollution heat maps resulting from a combination of air quality sensor data and additional data, such as meteorological and traffic information. A typical use of this type of maps can be observed in BreezoMeter [5]. By processing all collected data using proprietary algorithms, machine learning techniques, big data analytics, and air pollution dispersion modelling, BreezoMeter's team claims to be able to provide block level reliable air quality estimates at a 90% accuracy. Nevertheless, regardless of all the layered information used and sparse area covering, monitoring stations are still very much relied upon. When in the absence of these stations, an air quality estimate can hardly be calculated, even if every other layer contains information. For example, BreezoMeter does not contain information on Havana, Cuba, because there are no known installed air quality stations, even though there is satellite imagery of this city, as well as traffic data and weather forecasts.

In recent years, a rise in the use of LCS to measure air quality has been observed. PlumeLabs provides proprietary LCS, called Flow, to the public at a fraction of the price of traditional monitoring stations, although at the cost of measured pollutants diversity and their general accuracy. When activated, these sensors collect air quality data on $PM1$, $PM2.5$, $PM10$, NO_2, and volatile organic compounds every 60 seconds, which are then sent to PlumeLabs's database and used to create coloured street-by-street maps according to the registered AQI values. However, there are drawbacks to LCS, which are mainly related to accuracy and reliability. Recent studies have shown that results are not consistent between pollutants and different weather conditions, but acceptable nonetheless, showing high correlation when compared to official monitors [17], [18]. Although Flow already maps hundreds of big cities, the overall world-wide coverage is still far from complete. However, it does show promise, assuming more and more people acquire and use Flow.

One of the main issues with currently available air pollution mapping solutions, which is directly related to our work, is the lack of representation of individual pollution emission sources. For example, by centring BreezoMeter's map over the Bełchatów power station, in Poland, nothing indicates that it is one the most polluting factories in Europe, according to the European Environmental Agency [19]. Hence, being fully data-driven, existing solutions tend to fail in poorly sensor sampled areas. Our work overcomes this limitation in the state-of-the-art by allowing environmental experts to input their empirical knowledge into the mapping system, so that air quality maps can be generated even in the absence of sensory data.

Mapping solutions need to rely on a set of API to boost development, foster cross-platform operation, provide intuitive user interactivity, facilitate customisation, and to access multiple data providers. According to these criteria, three libraries standout as the most popular: Mapbox GL JS, OpenLayers, and Leaflet. These three rely on Open Street Map (OSM), which is a community-driven project focused on creating and maintaining open-source geographical data [20]. There are two commonly used map rendering approaches: raster tiles and vector tiles. The raster tiles approach downloads tiles at a fixed resolution and styling, which negatively affects interactivity, seeing that each time a change is made to the data set, the whole tile generation process must be redone. Contrarily, the vector tiles approach does all the rendering and styling at the client side, which means that changes can be made to the data set without having to recall assets. Hence, the vector tiles approach is the most adequate when map entity selection and dynamic styling are required, as it is the case of our tool. Leaflet does not support vector tiles unless an external plugin is included; and, even with the plugin installed, it requires a larger network bandwidth. Given that Mapbox GL is known to perform better than OpenLayers, even though by a small margin [21], it was selected for our tool. Furthermore, Mapbox GL implementation and documentation are more intuitive than the alternatives.

Empirical knowledge about air quality is introduced in our tool by environmental experts, by interactively associating air pollution simplified models to map entities (e.g., buildings, roads). This interactive process is considerably similar to the one often employed for image annotation, which, in turn, depends heavily on automatic/manual image segmentation. Image annotation is key for building large data sets for later use in supervised machine learning [22]. For instance, LabelMe [23] provides a web-based tool for easy labelling of images as well as a set of 10,000 images, 7,000 of which still to be annotated, and label information should include object classes, shape, locations, and other relevant labels. In the past two decades, plenty of research has been conducted on the topic of semi-automatic segmentation which is, to some extent, publicly available today in Microsoft's Paint3D or Adobe Photoshop's MagicWand, for example. Using the graph-cut optimisation technique [24], GrabCut [25] popularised semi-automatic image segmentation using a bounding box selection. After setting a bounding box around a certain object, a rough segmentation is automatically initiated to retrieve said object from the scene. The final interactive step is to draw scribbles over the object in order to obtain the desired result.

Despite all improvements observed in semi-supervised image segmentation techniques over the years, these may still fail in more challenging situations. To circumvent the challenges and limitations of current semi-automatic segmentation solutions and, thus, reduce the burden on the environmental expert, our tool relies on the semantic segmentation of city maps available in the OSM's database. Nevertheless, in some rare occasions, OSM fails to properly segment map entities (e.g., smaller buildings). To cope with these situations, the user is allowed to flexibly draw polygons and associate labels to them.

III. Tool Description

The tool was developed with two major goals in mind: to allow environmental experts to express their knowledge on how air pollution is correlated with city topology and time-varying factors; and to generate dynamic air pollution maps given the elicited expert's knowledge. Fig. 1 depicts the class diagram of the developed tool so as to cover these two major functionalities. The tool was developed in `node.js` [1] and thus can be run in any JavaScript compliant browser. The next sections discuss each of the classes and how the goals are realised.

A. Map Entities

One of the core structures of the tool is a map entity. This can represent a road, a building, or a green area. They

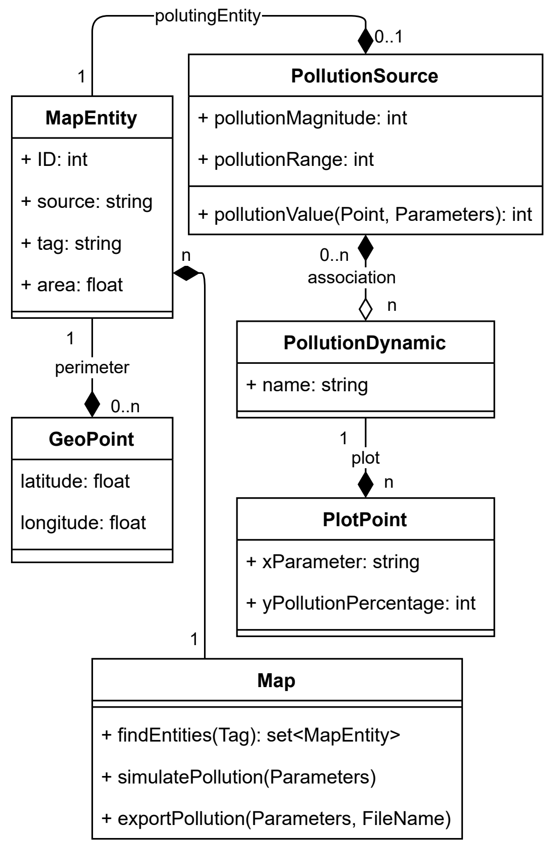

Fig. 1. Class diagram of the developed tool.

are created for every OSM object that is imported once the geographical area to be mapped has been selected by the user. The user can also create a map entity if it is missing from the OSM database. An important relation associated to each entity is its perimeter, as it is in a set of points evenly sampled on the perimeter that pollution is assumed to be emitted by the entity.

Figure 2 shows a screenshot of the tool with a map entity selected (left side of the image) and its properties displayed on the right panel. There are other map entities displayed with different colours depending on their tags: green for vegetation and blue for buildings.

In order to facilitate getting a specific entity, the tool provides a search functionality where the user can look for map entities by tag. This is useful when there is a need to assign a pollution source to multiple entities.

B. Pollution Heat Map

Pollution is represented as a heat map using the AQI colours (see Table I). Fig. 3 shows an example of a single heat map. All pollution magnitude values use the AQI scale, therefore the 250 value, used as magnitude in Fig. 3, should be interpreted as very unhealthy.

As map entities can be represented by polygons with various shapes or multiple lines, several points are sampled either on the perimeter of the polygon or on the line. Pollution is then

[1] `https://nodejs.org`

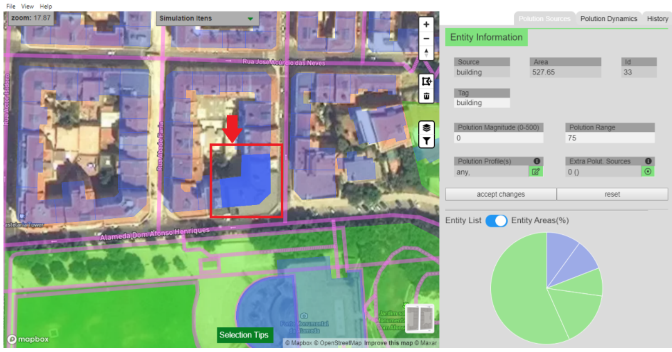

Fig. 2. After selecting a building (see red square), its bounding polygon becomes highlighted and its properties are displayed at the *Entity Information* panel.

Fig. 3. Pollution heat map after adding 250 (AQI units) to pollution magnitude and 200 pollution range to a single entity (building). The heat-point cluster dissipation is demonstrated by the transition between the colour red to green.

drawn as a circular heat map centred in each of these points. The amplitude and decay of these maps are defined by the user for each map entity in order to better represent what they believe to be the entity's typical pollution emission. All these circular heat maps are superposed to obtain the final heat map. The resulting visual scene can be seen in Fig. 4.

C. Pollution Profile

Any map entity can be associated with a pollution source. The latter is represented by class PollutionSource which only specifies a pollution magnitude and range, but these attributes only allow the expert to express a baseline pollution. The classes PollutionDynamic and PlotPoint allow the expert to represent pollution that depends on a set of parameters. The expert can create a pollution profile function that given a parameter value (day of week, hour of the day, wind main orientation) returns a percentage that is multiplied to the pollution magnitude. An example of such pollution profile is shown in Fig. 5.

As can be seen in the class diagram in Fig. 1, the user can assign a PollutionSource to any map entity, and the user can also create any number of PollutionDynamics as desired. It is only when the user creates an association between the two previous classes that an air pollution source is created. Moreover, a PollutionSource can have several pollution dynamics to represent different factors that influence a pollution source.

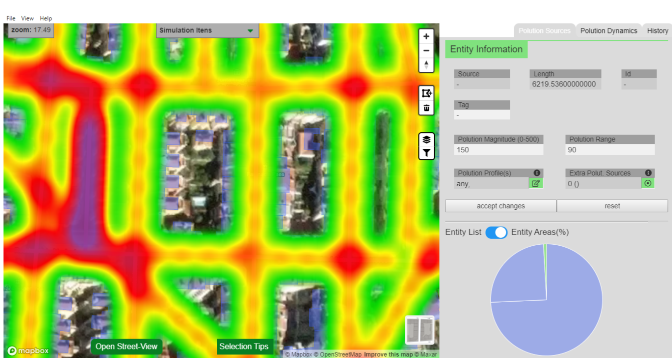

Fig. 4. Pollution heat map after adding 150 (AQI units) to pollution magnitude and 90 to pollution range to all visible roads.

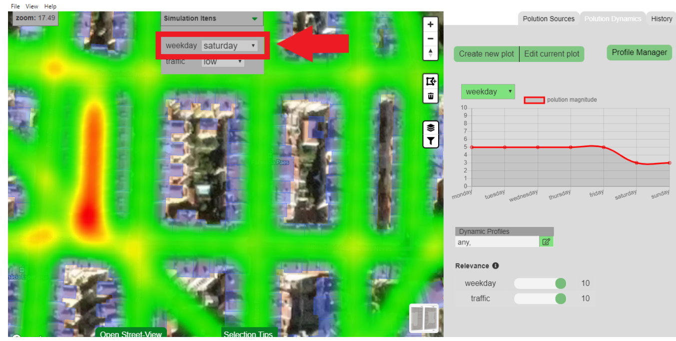

Fig. 5. Pollution profile that depends on day of week. Pollution is lower on weekends compared to weekdays. On the left panel, the highlighted box shows where the expert can select a set of parameter values.

D. Pollution Analysis

Once pollution plots are constructed, the expert can analyse different pollution scenarios. The expert can select a set of parameters, and visualise the pollution on the map. This functionality is represented by method simulatePollution in class Map.

An important feature is the ability to export simulated pollution data so that the user can perform further analysis. Method exportPollution allows the user to create a CSV file with a pollution given a set of parameters. Examples of further analysis are the ability to perform classification on the data created by the user.

By being able to construct pollution plots that depend on user defined parameters, the expert can represent knowledge on how a specific parameter will impact the pollution emitted by any pollution source that is associated. The expert can select a combination of parameter values and the tool displays a heat map representing the pollution, as can be seen in Fig. 5. Combined with the ability to export pollution data, the goals mentioned earlier in this section can be achieved with the pollution plot feature.

IV. Evaluation

The tool has been validated across three phases of development: requirements analysis, interface interactions, interface evaluation. In this paper we will focus extensively on the last phase, as it provided an usability score.

The set of requirements devised for the tool were analysed and revised with interviews with environmental experts. This consisted in creating a set of sketches to illustrate the main functionalities of the tool: the ability to select a geographical element, to create a pollution profile, and to assign a pollution profile to a geographical element.

Then, the Cognitive Work Analysis (CWA) framework [26] was used to plan the interface's structure beforehand. By employing this methodology, a constraint-based analysis of the interface was conducted in order to contain possible user interactions inside set boundaries. CWA's multiple phases facilitated early layout planning on what activities could be conducted given user's knowledge.

Despite being environmental experts, end-users are not expected to be computer-savvy. Therefore, in order to reduce interface usability problems, heuristic evaluations [27] were performed to guide interface adjustments across development iterations. Moreover, formative tests with six people (average age of 20.7) were conducted in order to identify application errors and bugs, as well as evident interface design issues. Each formative test began with participants just exploring the interface and becoming familiar with it. After exploring, a set of three exercises were presented, given a certain map location to work in: generate a heat map by associating pollution magnitude values to map entities; associate edited entities to profile *any*; and represent weekly variations using a chart. In order to identify poorer design decisions during the initial implementation phase, it is common practice to run formative tests with a small group of people, usually five.

Finally, the improved (current) version of the tool was subject to a summative evaluation, which is described in the following sections.

TABLE II

#	Task Description
1	Associating a pollution magnitude value of 100 (AQI units) to a specified road
2	Select and associate a pollution magnitude value of 150 (AQI units) to every building alongside a specified road
3	Represent weeklong variation in the viewport considering that at the weekend pollution is reduced to half
4	Select and associate a pollution magnitude value of 100 (AQI units) to all roads currently in the view-port
5	Associate a pollution magnitude of 500 (AQI units) to a specified non-segmented building
6	Represent daylong variation of the previous task's building, considering that this variation has half the impact of the weeklong variation

TABLE III

#	SUS Question Description
1	I think that i would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.
11	In the last task, the tool allowed me to create a pollution map given a zone and its properties.
12	In the last task, the tool allowed me to express my knowledge on pollution emission given a zone and its properties.
13	The pollution map I created given a zone, and its properties, is representative of what I expect is correct.

A. Summative Evaluation

Thirty participants were invited to test the tool, with the goal of assessing its interface efficiency and usability, validating the viability of the proposed knowledge-based approach for pollution mapping, and collecting feedback and improvement suggestions. Out of the 30 participants, a total of 6 are environmental experts (researchers and MSc students in environmental sciences, the latter being surrogates of established researchers), the other 24 users being non-experts (70% male and 30% female). All non-experts are MSc and BSc students, 3 from economics and management and the remaining from the computing area. Although the test results obtained with all participants were considered to assess the usability aspect of the interface, only the test results obtained with environmental experts were considered to assess the viability of the tool as a pollution knowledge elicitation tool.

Each user testing session started with a brief description of the project and the testing process itself. The participant was then asked to think out loud throughout the testing session, so it could be evident which interface aspect was being focused on, and which train of thought being followed when facing any potential challenge. After this introduction, a tutorial consisting of seven slides was shown to the participant, each slide describing a functionality of the application by displaying a video of its use, along with a short textual description. Afterwards, the actual user testing session begun. Three different phases composed this stage, the first two being interaction-oriented, whereas the third being opinion-oriented. Quantitative data such as completion rate or usability score were obtained in all three phases, while qualitative data was obtained when a participant required help to complete a task and in the final debriefing. Firstly, a set of six tasks with specific goals and related to main functionalities to be achieved using the interface was presented to the participant. The description of these tasks is shown on table II.

Secondly, a separate task requiring interaction with the application in order to produce a pollution map from scratch in a new location was presented. Finally, the participant was asked to fill the user interface evaluation questionnaire, System Usability Scale (SUS) [28], [29], consisting of 10 items with a five-point Likert scale ranging from *strongly disagree* to *Strongly Agree*, to assess the application's usability (see items 1-10 in Table III). Furthermore, three items were added to the questionnaire in order to further evaluate the viability of the proposed knowledge-based approach for pollution mapping (see items 11-13 in Table III).

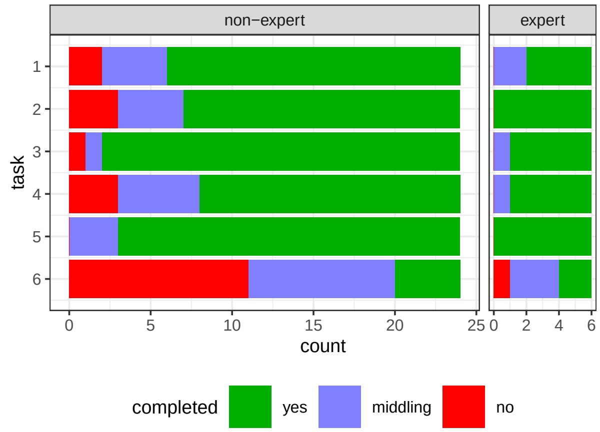

Fig. 6. Number of testers (count) that completed, partially completed, and did not complete each task. Testers are split in (environmental) experts and non-experts. Description of each task is presented in Table II.

B. Results

Overall, the summative evaluation, described in the previous section, shows that the tool's usability is satisfactory. Fig. 6 shows the number of testers that completed, partially completed, and did not complete task. The middle category was introduced to cope with the fact that in a couple of tasks there were testers that took over than five minutes to complete a task and/or required some kind of hint by the researcher conducting the test to be concluded. Overall, looking at the results presented in the figure, all tasks, except the sixth (chart creation and profile managing), were carried out successfully within the allotted time and without help by at least 70% testers. Generally, each task has a single goal that needs to be achieved for it to be considered successful.

Considering the first five tasks, a few usability issues were highlighted: the roads selection lines are too thin, requiring the user to be very precise when clicking on them; a few missteps were noticed in the multi-selection task which included the accidental selection of unwanted entities; the drop down menu for switching between simulation scenarios (corresponding to graphs items) was barely used; after drawing a polygon to create a new entity, users felt confused when trying to choose a tag to identify the latter.

The sixth task involved three implicit sub-tasks, that should be uncovered by the user (see Table IV): (1) to create a new pollution variation chart; (2) to adjust its relevance to half; and (3) to associate it to the same profile as the map entity created in the previous task. This task was considered to be successfully executed if all three sub-tasks were successfully executed. A middling success was assigned if only two sub-tasks were successful, and a failure in case of one or none. Table IV) shows that graph creation was successfully handled by the majority of users. The graph's relevance value defines how much it affects the associated entities' pollution emission. Table IV) shows that the sub-task associated to the adjustment of this parameter produced mixed results. Possibly, a

TABLE IV
SUCCESS RATE OF TASK SIX'S SUB-TASKS.

Sub-Task	Success
To create a new pollution variation chart	76%
To adjust the chart's relevance to half	43%
To associate the chart to profile of previous task	33%

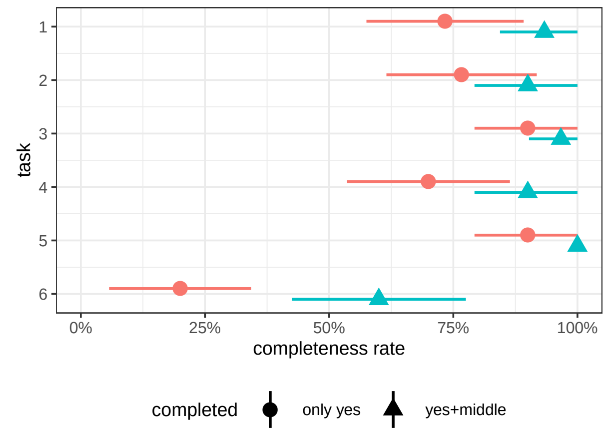

Fig. 7. Confidence intervals for completeness rate of tasks.

comprehensive reference in the tutorial would have decreased this issue's occurrence rate. Profile management proved to be the most misunderstood functionality, as shown in Table IV. User feedback obtained in debriefing revealed that the tutorial description and the accompanying video could have been more clarifying. Furthermore, once the interface element associated to profile management is opened, only one profile is visible: any. Perhaps the inclusion of more pre-defined profiles could help the user in obtaining a better understanding of the functionality.

To extrapolate the ability of any person to perform the tasks that we have analysed, Fig. 7 shows the 95% confidence intervals of the completeness rate. We have used the data both from experts and non-experts. We considered two types of completion, one where we only consider if a task was completed under the allotted time, and a second where we also consider the data were a person required extra time, or assistance by the researcher. Overall we can see that the confidence interval for tasks 3 and 5 is above 75% independently of how the tasks were completed. If we also consider the second type of completion, then the confidence interval of tasks 1, 2, and 4 raises above 75%. The confidence interval of task 6 is the only one that is low, which explains the feedback that we discussed in the previous paragraph.

A second phase of the summative evaluation process was carried out to assess the viability of the proposed knowledge-based approach for air pollution mapping. This builds upon the ability of environmental experts to use our tool to represent their knowledge regarding air pollution emission and diffusion, given satellite imagery. The specificity of this analysis led us to focus this second phase on the environmental experts.

These participants were asked to create an air pollution heat map to the best of their knowledge, given a pre-selected area in the corresponding satellite imagery. A proportion of 53% of the participants, which tried to create a pollution map in this task, wound up selecting multiple entities, mainly roads, and then associating the same pollution magnitude to all of them, whereas only 13% tried to introduce different values depending on the type and apparent width of the road. However, all participants exhibited difficulties in associating aggregate AQI values to entities, as they are used to perform more specific pollutant-wise assessments. As a result, the general opinion reported by the experts is that the tool should allow them to express their knowledge at the level of individual pollutants, instead of solely aggregate AQI values. Consequently, although experts recognised the novelty and value of the tool to their field, they displayed some difficulties in using it as a means to express their knowledge in its current form. Fortunately, their rich feedback generated a list of improvements that will be included in future versions of the tool. Among this feedback, we highlight the suggestion for using the tool in educational contexts.

To obtain more statistical information about the tool's usability, in a third phase of the summative evaluation, a usability questionnaire consisting of 13 items was presented to the testers. Ten of these items are from SUS, whereas the other 3 were appended so as to handle the specificities of task 6. Fig. 8 shows that, considering the first 10 item questionnaire, the obtained results are satisfactory. To better evaluate the results, a scoring method [29] was executed. The final score of the used method ranges between 0 and 100. Studies in usability [28] suggest that an acceptable interface should have a score of, at least, 68. The score 68.83 achieved by our tool indicates that the usability of its interface is satisfactory, although continued improvement should be implemented. If we consider the usability and learning dimensions in separate [30], our tool obtained a score of 68.13 and 71.67, respectively. This means that the usability of our tool is still satisfactory. The three last questions (non-SUS) of the questionnaire related to task 6 showed very different results from a usability standpoint, performing much worse than the previous ten, which is not surprising, considering the low success rate of the task. Returned feedback suggests that building upon and improving the developed tool by altering existing functionalities (e.g., individual pollutant definition) and adding new ones (e.g., including more data sources, and entity searching by tag) should produce better viability results. Information such as this, will be extremely useful in development of future iterations of the tool herein presented.

V. Conclusions

A graphical tool was presented with the novel idea of allowing users, namely environmental experts, to express and share their knowledge on air pollution emission and dispersion on an interactive map. This tool contributes with a novel way of complementing existing sensor-based air pollution mapping approaches, the ultimate goal being to use direct

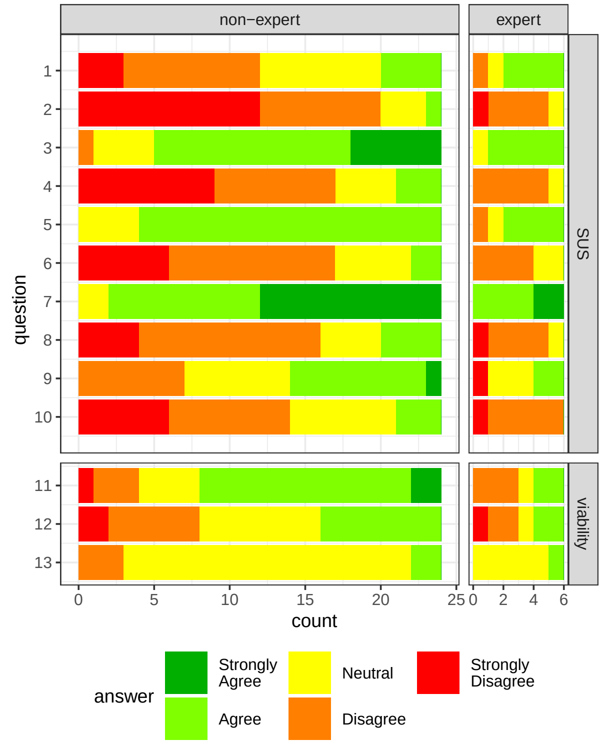

Fig. 8. Questionnaire results. Description of each question is shown in Table III.

expert knowledge to extend the range and efficacy of air quality predictions, particularly regarding individual pollution emitters. The obtained results in the evaluation phase demonstrate good usability levels. Participants generally found the interface to be intuitive. Functionalities such as entity selection, multi-selection, and polygon drawing were quickly recognizable. In some cases, adding pollution to entities and chart interactions required use of the tutorial, after which were easily realized. Some functionalities were not as easily used though, such as managing profiles, where feedback suggested that the corresponding tutorial slide should have been more clarifying.

While the tasks participants where asked to perform can be considered as specific to an environmental expert, they are rather focused on using the tool to perform an action which is not related to a real context. For instance in task 1, we ask the user to select a graphical element in the interface and assign a property to that element. This action is also found in another applications. If the task would be assign a high pollution value to the most road in a city, then this would require knowledge about the surrounding environment.

When faced with the idea underlying the tool, experts recognised its value and novelty. When requested to express their knowledge with the tool, a few improvements for forthcoming tool's versions became evident. The elements pointed out by the experts as the most valuable to bring the tool to the next level are: to allow the user to associate specific pollutant magnitude values to map entities, rather than a single aggregate AQI value; to provide the user with additional topological information, such as road width and buildings height, as these are known to greatly influence air dynamics; and provide the user with traffic information as, again, is key to predict air pollution magnitude. A reviewer also suggested including a scale with the AQI level.

We believe that future iterations of this tool have the potential of drastically improving it in terms of usability and viability. After resolving all usability issues and adding required functionalities, research should be made on additional data sources that provide some of the aforementioned information. Mapbox's own traffic layer, would be an option to consider for its relatively high coverage. Furthermore, having the option to view the map with 3D rendered buildings might help users to better visualise it and identify street canyons, which generally increase pollution. Again, Mapbox provides a layer that does this, although limited in the number of cities with complete information. The waqi.info API [16] overlays the map with all accepted air quality monitors and respective results. By including this API in the application, a new study could be conducted in which expert knowledge was compared with curated sensor data. This could answer the question of how relevant direct expert knowledge would be in pollution mapping. We also intend to study the value of the tool for expert knowledge-based filtering of noisy sensory data. Future versions of the tool will support the user when associating air pollution to map entities with semi-supervised segmentation techniques. Finally, we will study how this tool can be used in educational contexts. In a possible educational use case, a teacher could create an air pollution map to demonstrate features that cause or contribute to mitigate pollution. Students would then learn how to improve air quality.

References

[1] Y.-C. Lin, W.-J. Chi, and Y.-Q. Lin, "The improvement of spatial-temporal resolution of pm2.5 estimation based on micro-air quality sensors by using data fusion technique," *Environment International*, vol. 134, p. 105305, 2020.

[2] T. Becnel, K. Tingey, J. Whitaker, T. Sayahi, K. Lê, P. Goffin, A. Butterfield, K. Kelly, and P.-E. Gaillardon, "A distributed low-cost pollution monitoring platform," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 738–10 748, 2019.

[3] D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, T. Arn, J. Beutel, and L. Thiele, "Deriving high-resolution urban air pollution maps using mobile sensor nodes," *Pervasive and Mobile Computing*, vol. 16, pp. 268–285, 2015, selected Papers from the Twelfth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom 2014).

[4] L. Morawska, P. K. Thai, X. Liu, A. Asumadu-Sakyi, G. Ayoko, A. Bartonova, A. Bedini, F. Chai, B. Christensen, M. Dunbabin, J. Gao, G. S. Hagler, R. Jayaratne, P. Kumar, A. K. Lau, P. K. Louie, M. Mazaheri, Z. Ning, N. Motta, B. Mullins, M. M. Rahman, Z. Ristovski, M. Shafiei, D. Tjondronegoro, D. Westerdahl, and R. Williams, "Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?" *Environment International*, vol. 116, pp. 286–299, 2018.

[5] Breezometer. (2020) Global air quality map. [Online]. Available: https://breezometer.com/air-quality-map

[6] PlumeLabs. (2020) Real-time map of world pollution, by Plume Labs. [Online]. Available: https://air.plumelabs.com/air-quality-map

[7] IQAir. (2021) Live animated air quality map (AQI, PM2.5...). [Online]. Available: https://iqair.com/air-quality-map

[8] A. Lakhani, R. Balasubramanian, and B. R. Gurjar, "Air pollution monitoring and source characterization," in *Air Pollution, Health and Environmental Impacts*, B. R. Gurjar, L. T. Molina, and C. S. Ojha, Eds. CRC Press, 2010, ch. 17, pp. 493–518.

[9] P. Santana, A. Almeida, P. Mariano, C. Correia, V. Martins, and S. M. Almeida, "Air quality mapping and visualisation: An affordable solution based on a vehicle-mounted sensor network," *Journal of Cleaner Production*, vol. 315, p. 128194, 2021.

[10] B. Teles, P. Mariano, and P. Santana, "Game-like 3d visualisation of air quality data," *Multimodal Technologies and Interaction*, vol. 4, no. 3, 2020.

[11] A. C. Michalos, Ed., *Air Quality Index (AQI)*. Dordrecht: Springer Netherlands, 2014, pp. 120–120.

[12] S. Bumm, "Using satellite images to determine aqi values in california," 2019, course project. [Online]. Available: http://cs229.stanford.edu/proj2019spr/report/22.pdf

[13] M. B. Karuppasamy, S. Seshachalam, U. Natesan, R. Ayyamperumal, S. Karuppannan, G. Gopalakrishnan, and N. Nazir, "Air pollution improvement and mortality rate during covid-19 pandemic in india: global intersectional study," *Air Quality, Atmosphere & Health*, vol. 13, no. 11, pp. 1375–1384, 11 2020.

[14] A. Kumar and P. Goyal, "Forecasting of daily air quality index in delhi," *Science of The Total Environment*, vol. 409, no. 24, pp. 5517–5523, 2011.

[15] S. Selvam, P. Muthukumar, S. Venkatramanan, P. Roy, K. Manikanda Bharath, and K. Jesuraja, "Sars-cov-2 pandemic lockdown: Effects on air quality in the industrialized gujarat state of india," *Science of The Total Environment*, vol. 737, p. 140391, 2020.

[16] waqi.info. (2021) World's air pollution: Real-time air quality index. [Online]. Available: https://waqi.info

[17] N. Castell, F. R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. Broday, and A. Bartonova, "Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?" *Environment International*, vol. 99, pp. 293–302, 2017.

[18] S. Munir, M. Mayfield, D. Coca, S. A. Jubb, and O. Osammor, "Analysing the performance of low-cost air quality sensors, their drivers, relative benefits and calibration in cities—a case study in sheffield," *Environmental Monitoring and Assessment*, vol. 191, no. 2, p. 94, 1 2019.

[19] European Environmental Agency. (2017) Coal-fired power plants remain top industrial polluters in europe. [Online]. Available: https://www.eea.europa.eu/highlights/coal-fired-power-plants-remain

[20] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, 2008.

[21] R. Netek, J. Masopust, F. Pavlicek, and V. Pechanec, "Performance testing on vector vs. raster map tiles—comparative study on load metrics," *ISPRS International Journal of Geo-Information*, vol. 9, no. 2, 2020.

[22] J. Weston, S. Bengio, and N. Usunier, "Wsabie: Scaling up to large vocabulary image annotation," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, T. Walsh, Ed. AAAI Press, 2011, pp. 2764–2770.

[23] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "Labelme: A database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1, pp. 157–173, 5 2008.

[24] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary region segmentation of objects in n-d images," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1, 2001, pp. 105–112.

[25] C. Rother, V. Kolmogorov, and A. Blake, ""grabcut": Interactive foreground extraction using iterated graph cuts," in *ACM SIGGRAPH 2004 Papers*, ser. SIGGRAPH '04. New York, NY, USA: Association for Computing Machinery, 2004, p. 309–314.

[26] K. J. Vicente, *Cognitive Work Analysis*. CRC Press, 1999.

[27] J. Nielsen, "Enhancing the explanatory power of usability heuristics," in *CHI '94: Conference Companion on Human Factors in Computing Systems*, 1994, pp. 152–158.

[28] J. Sauro. (2011) Measuring usability with the system usability scale (sus). [Online]. Available: https://measuringu.com/sus

[29] J. R. Lewis and J. Sauro, "Item benchmarks for the system usability scale," *Journal of Usability Studies*, vol. 13, no. 3, pp. 158–167, 5 2018.

[30] ——, "The factor structure of the system usability scale," in *Human Centered Design*, M. Kurosu, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 94–103.