THE KEY ROLE OF GEOGRAPHIC INFORMATION IN EXPOSOMICS: THE EXAMPLE OF THE H2020 PULSE PROJECT

D. Pala [1, *], L. Annovazzi-Lodi [2], R. Bellazzi [3], N. Fiscante [4], M. Franzini [2], C. Larizza [3], A. Pogliaghi [4], L. Raso [4], M. T. Rocca [2], F. Sapio [4], V.Casella [2]

[1] Dept. of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy – daniele.pala02@universitadipavia.it
[2] Dept. of Civil Engineering and Architecture, University of Pavia, Italy – laura.annovazzilodi01@universitadipavia.it, (marica.franzini, maricateresa.rocca, vittorio.casella)@unipv.it
[3] Dept. of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy – (riccardo.bellazzi, cristiana.larizza)@unipv.it
[4] GeneGIS GI Srl, Milano, Italy – (a.pogliaghi, n.fiscante, l.raso, f.sapio)@genegis.net

Commission IV, WG IV/4

KEY WORDS: Exposomics, WebGIS, Public Health, Spatial Enablement, Asthma, Diabetes

ABSTRACT:

Exposomics is a novel concept that indicates the combination of all the external factors we are exposed to throughout our entire life, as the environment we live in, our lifestyle and behavior are able to have a notable influence on our health. The quantity and typology of environmental factors we are exposed to are clearly dependent on the geographical location of each individual, e.g. some areas are more polluted that others and even the social characteristics of a certain place can have an effect on the way we behave, exposing us to different levels of risk of developing certain diseases or exacerbating existing ones. In this context, the PULSE project, briefly described in this paper, is building an advanced system to identify the effect of a complex set of environmental and social exposures in the big cities, that represent the most complicated environment from this point of view, and mitigate health risk related to common diseases such as asthma, type 2 diabetes and cardiovascular diseases. This system is composed by several parts, most of which apply advanced spatial analytics and geographic information-based tools to estimate health risk in a precise way, providing both citizens and public health officers with tools to monitor it. This paper summarizes the work performed in the project using these analytics, and quickly describes some of the tools in which geographic information has been applied in the most innovative way.

1. INTRODUCTION

The percentage of the world's population living in urban areas is projected to increase in the next decades (Organization and UN-Habitat 2016). Big cities are heterogeneous environments in which socioeconomic and environmental differences among neighborhoods are often very pronounced, leading to a new burden of problematics to be faced by the public health authorities related to air pollution, wellbeing, social discrepancies and quality of life. Several studies have demonstrated that due to an increase in air pollution and a change in lifestyle, several conditions such as asthma, type 2 diabetes and cardiovascular diseases are becoming more common every year (Anandan et al. 2010). These diseases share a common multifactorial characteristic, i.e. they are the results of a complex summation of different environmental exposures, often difficult to isolate. To face these problems, the international project named PULSE (*Participatory Urban Living for Sustainable Environments*), funded by the European commission, partners with 7 cities worldwide to create a collaborative system that involves directly both citizens and public health authorities to prevent the development and/or complications of these diseases in the urban environment. PULSE works inside the exposomics concept, i.e. the study of all the combining environmental factors human beings interact with and of how they influence health and wellbeing.

The PULSE system is made of a complex multi-technological architecture composed by several parts that connects citizens and public health officers in a continuous data exchange that allows a prompt response to health risk exposure.

Spatial Enablement, i.e. the addition of a geographic description to a set of data, is one of the most fundamental concepts in PULSE. It has been applied to a vast variety of data of all typologies, such as environmental (pollution, satellite LST temperature maps, vegetation etc.), socioeconomic (poverty rate, land use, education, obesity rate, smoking rate etc.) and demographic (age, race, sex etc.), in order to build a comprehensive exposomics study where health risk is the result of a combination of multiple factors that are strictly dependent on geography. In detail, PULSE features a modern and vast WebGIS that has a fundamental role in the innovative paradigm of the project, since it represents the integration of the geographic analytics in health outcomes, urban planning and socioeconomic analyses. The WebGIS is a powerful visualization tool, that features a set of innovative functions such as a temporal bar to navigate through past data with a daily resolution and side by side visualization of different maps, that allows to visually inspect the possible correlations between different phenomena (e.g. health outcomes and environmental factors) at the same time. It contains several layers about socioeconomic data, pollution records, health indicators (prevalence and incidence of diseases, hospitalizations etc.), satellite imagery (NDVI index maps, LST temperature maps), basic demographics and urban planning indicators. Besides, the PULSE WebGIS is also concepted as an active analysis tool, since spatial descripted data are being used to predict health-related phenomena in the urban environments at a neighborhood level. For example, a study has been carried out in New York City in which it was found that asthma hospitalizations can be predicted using a Geographically Weighted Regression (GWR)

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-283-2020

model that integrates maps on pollution, demographic factors and socioeconomic indicators such as poverty rate, land use and education level. This model showed that all the different factors' influence on the outcome has a strong dependency on space, since different models describe the data in different ways depending on the neighborhoods they are applied to, thus enlightening the importance to study public health issues at a neighborhood level in order to plan effective interventions, reducing costs and increasing accuracy. Another study, performed integrating satellite images of NYC with high spatial resolution health data, found that the percentage of green areas within the city is correlated to a lot of health outcomes.

Finally, spatial enablement in exposomics is used also in an interactive way in two tools that are integrated inside the PULSE dashboard (see next section): an agent-based simulation tool ad a pollution personal exposure calculator. The first one is a tool that implements the relationships found among asthma hospitalizations and a variety of different factors in a predictive way, allowing the public health operators to simulate how and if a specific set of interventions can influence the probability of hospitalizations in a certain environment. The simulation is set on real GIS data of different neighborhoods of New York and follows an agent-based paradigm, where agents (people in this case) interact with the environment and with each other following specific rules. The second one is a personal exposure calculator, developed for the urban area of Pavia, Italy, that applies state-of-the-art low-cost sensor technology and positioning to estimate a personalized pollution exposure for each user that gives consent to be tracked by the GPS. Thanks to a dense sensors network that provides real-time measurements, a homogeneous pollution map can be created through interpolation, and using global positioning technology it is possible to compute the exact quantity of pollution a user is exposed to depending on their spatial position and intensity of physical activity, which the actual air intake depends on.

In the next sections of this paper, a more detailed descriptions of these tools is presented, showing how GIS tools, satellite imagery, positioning and spatially enabled big data algorithms are used in a proactive way to create personalized risk calculators and tools to ease the public health authorities' intervention strategies to improve health and quality of life in the urban environments of the new era.

2. GEOGRAPHIC INFORMATION IN PULSE

2.1 The PULSE System and WebGIS

The multi-technological platform developed in PULSE is composed by different interconnected parts, some of them addressed to the citizens, some of them specific for policy makers and other for both. In detail, these parts are:

- A personal App for the users, through which they can send their own data and receive personalized feedback on their health risk in return.
- A WebGIS that shows on maps all the spatially described data of the PULSE database, that features several tools to overlap, compare and navigate data in space and time.
- A dashboard for the Public Health Authorities, that can be used to inspect aggregated data about their city and spot critical situations that need interventions, and to simulate scenarios based on the real data acquired in the city.
- A complex back-end infrastructure, that connects all the technical parts and in which the datasets reside.

The WebGIS is one of the most important and innovative parts of the system, since it provides a spatial representation of the data stored in the database, allowing to monitor different kinds of public health phenomena and explore their location-related dynamics. The WebGIS stores and represents data of different kinds coming from all the 7 test sites of the project: Barcelona, Birmingham, Keelung, New York, Paris, Pavia, Singapore. Some examples of this data are:

- Pollution maps, showing present and past data coming both from official monitoring stations and sensors acquired locally within the project;
- Satellite images, such as LST temperature images, showing other pollution data, NDVI index maps, vegetation maps etc.;
- Socioeconomic and urban data, such as poverty rate, crime rate, recycling rate etc., with their intraurban differences;
- Demographic data, showing how the population differs in age, sex, race, and density within the city;
- Health data, such as prevalence and incidence of certain diseases, hospitalizations etc.

All these data can be inspected separately or unified in semitransparent layers that can be overlapped in order to quickly compare the distributions of different variables throughout the city. Given the large amount of data stored in PULSE's database and how it is considerably interconnected, some peculiar and original tools have been developed and added to the PULSE WebGIS. Specifically, these tools are:

- A temporal bar that allows to navigate data in time and visualize past data. For most of the cities, the same kind of data is available in different spatial and temporal granularities. Through the temporal bar, the users can explore past data and choose the proper temporal resolution to represent it;
- A side by side visualization tool, that allows to position different maps of the same city aside, making it easier to compare different variables and visually spot correlations between indicators.

Both health and socioeconomic data are usually spatially enabled, i.e. they can be described with measures and indicators that depend on the geographic position where the data is taken.

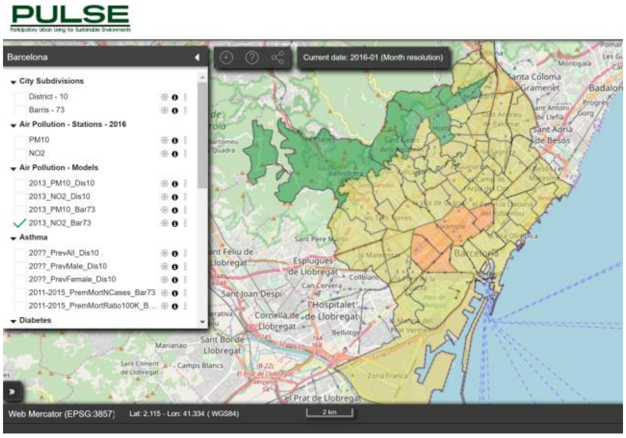


Figure 1. screenshot of the PULSE WebGIS page related to the city of Barcelona. This map, made partially transparent, shows the average quantity of NO2 measured in the different neighborhoods of the city in the year 2013

For example, considering the hospitalization rate of a certain disease, this rate can change depending on the considered hospital or on the zipcode where the patients live, depending on specific risk factors that can have differences within the city themselves. Spatial Enablement is one of the fundamental paradigms in PULSE, and it has been explored and applied in different systems, both for representation and analysis purposes. Although this concept is not new, there is a generalized lack of research concerning how spatial enabled public health indexes can vary within a city, since data collected at a high spatial granularity is usually absent or not sufficient for the creation of statistically significant analyses. Thanks to the creation of the PULSE system, that gathers large quantities of intraurban data, these difficulties have been partially overcome. The WebGIS is the visual example of this innovation.

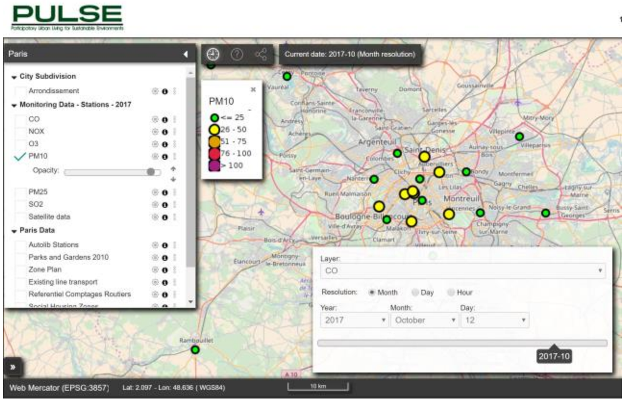


Figure 2. Another part of the PULSE WebGIS, that shows data coming from air quality sensors in the city of Paris. On the bottom right it is possible to see the temporal bar, that allows to navigate data in time and choose the temporal granularity

Figure 1 and Figure 2 show some snapshots of the PULSE WebGIS, enlightening how it allows to visualize different kinds of data with various spatial distributions and navigate it using advanced tools such as the temporal bar, visible in figure 2.

Among the environmental variables that are known to have an effect on air pollution and health risk in general, climate-related variables are also monitored with particular attention using satellite images. There are two kinds of maps which are related to climate monitoring and to mitigation of climate change effects which are available in the PULSE system:

- LST maps: Land Surface Temperature maps showing surface temperature in a continuous way; they are particularly important to study the heat waves and heat islands in cities;
- NDVI maps – The Normalized Difference Vegetation Index maps are a simple and powerful way to detect green areas in cities; it is well known that green areas have a strong mitigation power for several effects of climate change such as increased temperatures, heat waves, air pollution, string rainstorms.

Two screenshots are reported for the Pavia pilot, highlighting the integration of the mentioned maps within the PULSE WebGIS; they represent the LST (Figure 3) and the NDVI map (Figure 4), respectively.

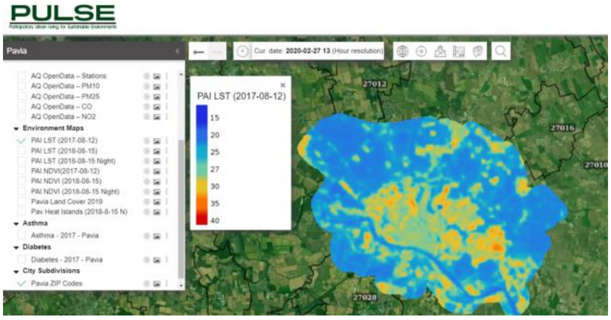


Figure 3. LST map for Pavia on August 8th, 2017

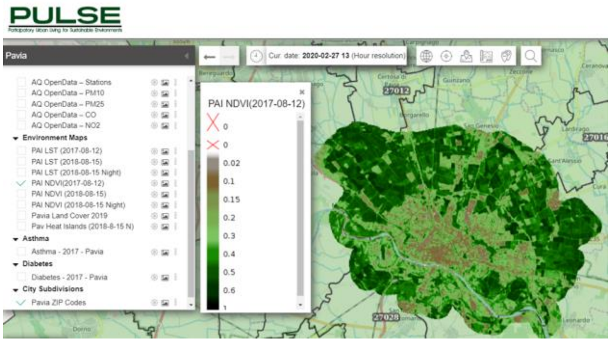


Figure 4. NDVI map for Pavia; green pixels have a high NDVI value and correspond to fields or gardens; grey areas have a low NDVI value and correspond to built-up areas

2.2 Geographically Weighted Regression

Among the different applications of spatial enablement, an interesting study has been performed in the context of PULSE, providing a demonstration of the necessity to model health with a very high spatial resolution on a big city and offering the basis for the creation of new useful tools and studies.

This study, published in 2019 (Pala et al. 2019), had the aim to model the relations among the asthma hospitalization rates in the different neighborhoods of New York City (NYC) and a combination of environmental, demographic and socioeconomic factors. After a brief exploratory analysis, performed with a spatial clustering algorithm, i.e. a clustering algorithm that weights cluster depending on their geographic proximity (Grubesic et al., 2014), the main method that was used was Geographically Weighted Regression (GWR) (Brunsdon et al., 1996), i.e. a method based on the creation of a dense network of linear regression models overlapped to a geographic area where covariates' values vary depending on their location. All the linear models created in this network are weighted with a parameter tuned in order to increase the probability of having similar relations in geographically close areas. In our study, we tested both univariate and multivariate models using the asthma hospitalizations rate as dependent variable and combining the effect of a large number of independent variables, such as air pollution, poverty rate, education, age, race, obesity rate, smoking rate, recycling rate, industrial land use. Most of the found relations were not new, but there were two main findings in this study: first of all, it appeared that in NYC socioeconomic status has a larger influence on asthma than bad air quality (poverty rate was one of the most significant variables), and, above all, the influence of each one of these variables or of the combined multivariate effect changes a lot depending on the neighborhoods in which the relations are studied, and urban public health cannot be addressed properly without using high spatial resolution data, as the heterogeneity of the urban environment and population can make the health situation change dramatically even in a very tight space. Therefore, considering the city as a unique environment can hide important local public health issues.

Figure 5 shows one of the results of this algorithm, i.e. the univariate model that relates asthma hospitalizations to poverty rate. The results show that the correlation is generally high, but presents several differences depending on geographical location.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-283-2020 | © Authors 2020. CC BY 4.0 License.

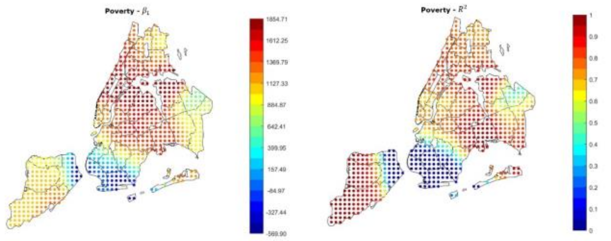


Figure 5. Result of the univariate model that correlates asthma hospitalizations to poverty rate. A different model is computed for a different geographical point of the city. A map of the β coefficients is shown on left, and a map of the corresponding R^2 scores is shown on the right

2.3 Deep Learning Analysis of Correlations between Urban Structure and Health Indexes

One of the most important parts of the PULSE platform is the so-called PHO dashboard, i.e. a comprehensive tool for the public health authorities that allows to analyze aggregated data about the city in order to spot the location and the magnitudes of public health criticalities and organize proper targeted interventions accordingly. The WegGIS and the GWR demonstrate how the use of geographic information can play a fundamental role in representing data and analyzing it in order to create predictive models, in addition, spatial enablement in PULSE is used also as an active tool for the optimization of the intervention strategy. A first example of this, is another study conducted in PULSE and published in April 2020 (Pala et al. 2020) on the journal Sensors, were we used a deep learning approach to analyze the relation between urban structure and several public health indexes. This study proposes an analysis pipeline to optimize the intervention design strategies following the idea that if there is a correlation between urban landscape and health indexes, then areas with similar urban structure will have a similar health situation and respond in a similar way to the same kind of interventions.

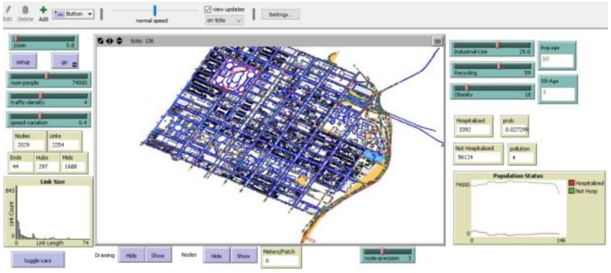
The pipeline that was developed starts with a simple high-resolution satellite image of New York City, coming from the National Agriculture Imagery Program (NAIP), that has been subdivided into smaller images representing parts of the city. These images have been then associated to 25 health indicators describing the general status of health of the population, the percentage of prevention activities and how people are taking care of their health. A Convolutional Neural Network was then selected to perform the embedding of the images after testing different models and using their capacity to group together images with a high percentage of green color as an evaluation. The embedded measures where then clustered using a K-means algorithm that found four different clusters that represented different urban structures, spacing from green parks to industrial or highly urbanized areas. Correlating these clusters with the health indicators through some proper statistical tests and a logistic regression, it was found that the distribution of most of the health indicators was highly dependent on the cluster the image was assigned to. This kind of analysis can serve as an aid to the intervention design process, since spotting similar areas, even if geographically distant, can reduce times and costs of the urban planning process, knowing here and how to intervene without studying the situation in each specific neighborhood.

2.4 Spatially Enabled Agent-Based Models for Urban Planning

As already said, spatial enablement and geographic description in PULSE are used in a predictive way, creating functionalities to facilitate intervention planning. The PULSE dashboard comprises a set of visualization tools to inspect aggregated data from the city and a set of simulation tools, that allow the operators to estimate how some indicators in the city would be affected by variations in some variables, modeling different possible scenarios, answering to "what-if" questions and simulating the effect of interventions. This tool is based on the Agent-Based modeling technique (ABM), a widely used modeling paradigm that is centered on the construction of complex systems made up by entities, called agents, that interact with each other and with the environment according to specific mathematical relations. Agents can represent a large variety of real-world entities, such as people, animals, plants, bacteria, viruses, cars, trucks, ships etc., therefore ABMs are widely used in many research fields, ranging from epidemiology to econometrics and social sciences.

The idea of using these models as a simulation tool for the public health operators can be shown with a key example of a model based on the GWR study described in section 2.2, where we found equations that describe the relation between asthma hospitalizations and a set of variables of diverse kind. The model has been developed in NetLogo, a specific open source software for ABM, that provides a large number of tools for modeling and agent representation. The models are discretized in time and each time step is named "tick". This model is set on East Harlem, one of the NYC neighborhoods with the highest hospitalization rates, and the simulation takes place on a background made of GIS data representing streets, sidewalks, buildings and parks. The user that runs the simulation can set several initial parameters that showed to be correlated to asthma hospitalizations risk and that can be possibly modified through some kind of intervention doable by a public health authority. In detail, these parameters are: traffic density and car speed variations, which control air pollution, obesity rate, percentage of territory used for industrial or commercial purposes, recycling rate. The observer can also set the initial age of the population, giving as input mean and standard deviation of a normal distribution. This allows to study two different effects: first, asthma risk depends on age, among other things, as we saw that hospitalizations are more common in people younger than 18 and in people older than 65 (with a smaller peak), and decrease to a minimum between 18 and 65 years of age; secondly, the age of the population allows to take into account also time inside the model, as people tends to age when the simulation is running. In detail, each tick of the model corresponds to six months, a reasonable quantity of time to see the effect of urban interventions that take some time to be applied. Furthermore, after each tick of the model every person has a certain probability to be hospitalized due to the combination of factors that contribute to the underlying GWR model, where the weights are set according to the results of the study reported in section 2.2.

After one tick, each hospitalized patient most likely recovers (99%), as our data showed that mortality for asthma is very low, and ages six months. After they are 75 years old, their probability of dying from other causes starts to increase every year. The control panel of this simulation model is visible in Figure 6.



zoom 0.8
num-people 74000
traffic-density 4
speed-variation 0.4
Nodes 2029
Links 2254
Ends 44
Hubs 297
Mids 1688
Meters/Patch 0
node-precision 3
Industrial-Use 29.0
Recycling 59
Obesity 18
Avg-age 10
SD-Age 3
Hospitalized 3392
prob 0.027299
Not Hospitalized 56124
pollution 4

Figure 6: interface of the East Harlem asthma hospitalizations agent-based model.

The same model can be built also for other neighborhoods and other cities. Extended versions of this model that include advanced traffic dynamics and human behavior dynamics are being developed, in order to increase the level of realism and accuracy in the predictions and to widen the range of possible interventions that can be simulated.

2.5 Environmental Personal Exposure Calculator

Spatial analytics and geographical tools are fundamental constituents in another PULSE system that serves as an aid to citizens and a health risk mitigator. One of the aims of PULSE is to increase accuracy and spatiotemporal granularity of environmental and health measurements in the urban areas, as this is often still insufficient to gather proper data that can be used to perform health risk analysis at a neighborhood level. For example, air quality is usually measured by a few monitoring stations spread in a large environment, controlled by local environmental agencies and public authorities. These monitoring stations are usually very accurate and also expensive and difficult to maintain; therefore, it would be hardly sustainable to deploy a large number of them in the same urban environment. As a consequence, local effects in the air pollution measurement far from the monitoring stations could be missed, as much as other local effects in the monitoring station's site could potentially disturb the official measurements. On the other hand, low-cost sensors are becoming more common, although they often show lower accuracies compared to the more expensive ones. Several studies are trying to find the perfect trade-off to maximize accuracy, cost reduction and spatial density all together (Özkaynak et al., 2013; Glasgow et al., 2016; Sanchez et al., 2019). In the development of PULSE, we studied several possibilities comparing different low-cost sensors that could be used to create dense networks and measure air pollution together with the official monitoring stations. A valid possibility was identified in the PurpleAir sensors (Purple Air website), that present relatively high quality and low price, and can be easily installed on balconies or private spaces, since they just need a wall to be drilled on, Wi-Fi connection and an electrical outlet. For a first experiment, we acquired several sensors and installed them in the city of Pavia, Italy, located in a geographical area that suffers from frequent air pollution problems, though agreements with the local city council and environmental agency. We tested some of the sensors through comparison with the official monitoring stations and found that the measurements had a high linear correlation, although with a small offset. Currently, there are 48 sensors deployed in the city that, together with the official monitoring stations, property of ARPA (the local environmental agency), create a dense interpolated network.

Thanks to this great number of sensors deployed in Pavia, an innovative personal pollution exposure calculator, that aims at reducing the effect of pollution on the population by following each person personally, is currently under development. Indeed, an advanced personal exposure functionality should have the following distinctive characteristics:

- to be open: being available to all the citizens and not requiring any special device;
- to be dynamic: having the capability to assess the instantaneous inhaled pollution and thus being able to sum up for any time frame chosen by the user (one minute or one hour) and for any time.
- being continuously available: it is performed routinely rather than for a special period when special arrangements are performed;
- being upgradable at no-cost for the user: in the case that more advanced methodologies become available for air quality monitoring, changes will be performed in the monitoring stations and in the processing equipment, supposed to be owned by the municipalities; users will only have to upgrade the App running in their mobile phone.

The system exploits two main parts of the technology developed in PULSE and its spatial enablement: the data coming from the dense sensor network deployed in Pavia, interpolated in space; the GPS tracking functionality of PULSE's app, that uses the geolocation feature of the users' phones.

At the present time, this system has been fully implemented only in the city of Pavia, but some tests are already performed for New York City and Barcelona; future plans are to extend it to the all the other cities participating in PULSE. Integration in the PULSE dashboard is also in progress.

In the personal exposure calculator, each user, after reading and agreeing to a detailed consent form, is followed all day long with the GPS tracking and his/her movements are used to draw a trajectory. Meanwhile data coming from the sensors are continuously stored and interpolated in a homogeneous map. These two elaborations are finally unified in an estimation of the cumulative intake of pollutants that were breathed by the use throughout the day or in a set amount of time. This operation is performed through an estimation of the number of breaths that were taken by the user in each minute of the day and the correspondent air volume intake, considering four different levels of physical activity: at rest, walking, running and driving a car. These quantities are reported in Table 1 (Breathe, 2016).

Speed [km/h]	Status	# breaths per minute	Air volume per breath [litre]
< 2	At rest	15	0.6
2 - 6	Walking	28	1.8
6 - 15	Running	40	2.5
> 15	Driving a car	15	0.8

Table 1. Air intake estimation (per breath) based on 4 levels intensity physical activity performed by the user

An interpolated pollution map is obtained applying the Gaussian kernel interpolation algorithm (Wilson and Nickisch, 2015) to the sensors' measurements in order to estimate the pollution level in every location of the city. Figures 7 and 8 show, respectively, an interpolated map of an air pollution measurement coming from all the PurpleAir sensors deployed in Pavia and a trajectory of a user in the city. Putting those two things together, it is possible to estimate the intake of pollution (Figure 9) that was taken by the user considering their position in each specific time frame.

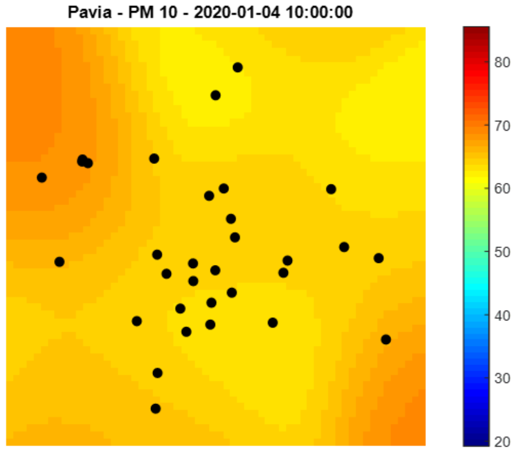


Figure 7. The interpolated continuous model for PM10; day selected is January 4th, 2020 and time shown is 10:00 AM

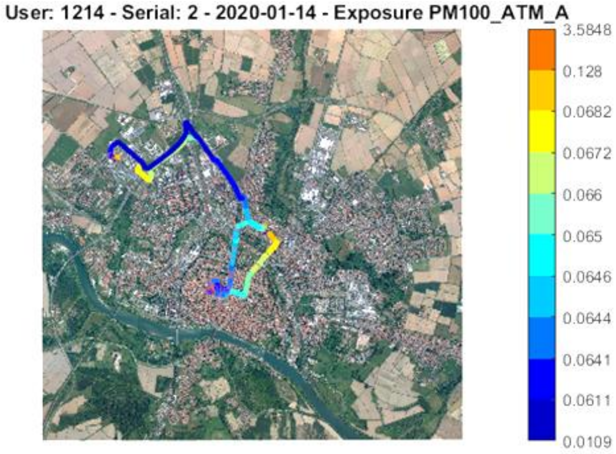
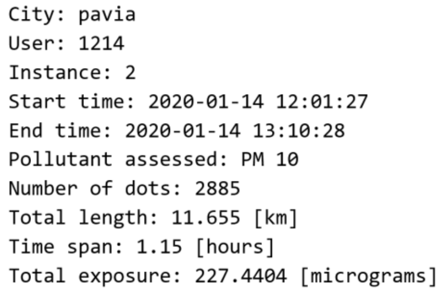


Figure 8. Personal exposure calculated for a user in Pavia on January 14th, 2020



```
City: pavia
User: 1214
Instance: 2
Start time: 2020-01-14 12:01:27
End time: 2020-01-14 13:10:28
Pollutant assessed: PM 10
Number of dots: 2885
Total length: 11.655 [km]
Time span: 1.15 [hours]
Total exposure: 227.4404 [micrograms]
```

Figure 9. Summary statistics

3. DISCUSSION AND CONCLUSIONS

The word “exposomics” refers to the study of all the environmental factors we are exposed to throughout our lives and their effect on our health, wellbeing, life expectancy and quality. This is a novel concept that derives from the awareness that epigenetic factors can influence our life more than we used to think. There are several difficulties when it comes to studying exposomics, since the number of exposures can be extremely high and difficult to identify, plus they can combine in different ways creating effects that are extremely hard to separate. Furthermore, environmental factors are strictly linked to geographic location, so exposomics studies cannot be applied accurately without proper geostatistics tools. With the technological progress of the last decades, that allows to collect and analyze quantities of data that are becoming larger every day, a growing number of studies are being conducted in this field. Among these, the PULSE project focuses on how exposures can be modeled and discovered inside the urban environments, that are characterized by high heterogeneity and varying environments in small distances. Big cities are dynamic environments in which most of the world population resides, and social and environmental contrasts are particularly emphasized. This leads to the necessity of performing studies with a high spatial resolution, since every neighborhood can be in a totally different situation compared to the rest of the city. Unfortunately, this is not an easy task, since collecting data at a sufficient spatial granularity to perform meaningful analyses at a neighborhood level is difficult and expensive, as it is performing adequate interventions in each limited geographical area.

To overcome, at least partially, these problems, PULSE proposes a Big Data approach, building a solid infrastructure composed by several different technologies that allow both to monitor every single participant personally and to intervene on the general situation of the city locally. To this end, the system is based on advanced geographic tools, adequately created to perform both analysis and visualization. The development of these tools start from the WebGIS, an advanced visualization tool that collects and shows all the data that can be geographically descripted and that can show how exposure factors and health outcomes are spread in a certain environment, with the aid of enhanced utilities such as the temporal bar and the side by side visualization option. In addition, geographical tools are used also to perform analysis and predictions, as it was done with the GWR, where we were able to develop a model that computes the risk of asthma hospitalizations according to the neighborhood of residence and the personal characteristics of each inhabitant. Results of this analysis can have several applications, not only to understand how health risk is linked to the different areas of the city, but also to predict how it could be mitigated in each neighborhood. To this end, the PULSE dashboard contains a set of visualization and simulation tools that can help to perform the right decisions having an idea of what are the specific problematics to be resolved and what could be the consequences of a certain intervention plan. This is mainly done by agent-based modeling technology and other specific analysis tools such as deep learning pipelines to analyze satellite images and create fast lanes for the organization of urban planning activities. Besides these aggregated tools for the policy makers, geospatial enablement is used also to aid the citizens directly, as demonstrated by the personal app and the personal exposure calculator, that uses geographical analytics unified to bioengineering models and dense sensing technology to assist each single user in reducing his/her health risk, or at least the component deriving from air pollution.

Of course, this approach is not immune to problematics, since each single city in the world has different data collection protocols, that make the creation of integrated systems and analysis pipelines very hard. Also, geographic data often comes with errors or missing values that are not properly treated, creating the necessity of long preprocessing and harmonization activities. As a result, most likely the tool developed for a city will not be suitable to be used in another city, as the creation of unified analysis and intervention protocols is still mostly impossible.

In conclusion, PULSE’s innovation is based on the application of high-level spatial analytics to a new concept of exposomics, where health risk is considered to be a combination of environmental exposures that cannot be separated from their geographical link. Technologies such as GIS tools, satellite imagery, positioning and spatially enabled big data algorithms are used in a proactive way to create tools that aim at assisting both citizens personally and public health authorities that can take informed public health decisions.

ACKNOWLEDGEMENTS

The authors wish to thank all the PULSE consortium and the collaborators to this project, specifically the New York Academy of Medicine, the College of Global Public Health at the New York University, the municipality of Pavia and the local environmental agency (ARPA Lombardia). This work has been funded by the European Commission Horizon 2020 Framework programme under grant agreement GA-727816.

REFERENCES

Breathe, 2016. Your lungs and exercise, vol. 12, no. 1, 97–100, doi.org/10.1183/20734735.ELF121.

Anandan, C., Nurmatov, U., Van Schayck, O.C.P., Sheikh, A., 2010. Is the prevalence of asthma declining? Systematic review of epidemiological studies. Allergy, 65(2), pp.152-167. https://doi.org/10.1111/j.1398-9995.2009.02244.x.

Brunsdon, C., Fotheringham, A.S. and Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical analysis, 28(4), pp.281-298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Glasgow, M.L., Rudra, C.B., Yoo, E.H., Demirbas, M., Merriman, J., Nayak, P., Crabtree-Ide, C., Szpiro, A.A., Rudra, A., Wactawski-Wende, J., Mu, L., 2016. Using smartphones to collect time–activity data for long-term personal-level air pollution exposure assessment. Journal of exposure science & environmental epidemiology, 26(4), 356-364. doi.org/10.1038/jes.2014.78

Grubesic, T.H., Wei, R., Murray, A.T., 2014. Spatial clustering overview and comparison: Accuracy, sensitivity, and computational expense. Annals of the Association of American Geographers, 104(6), 1134-1156. https://doi.org/10.1080/00045608.2014.958389.

Organization, World Health, UN-Habitat. 2016. Global Report on Urban Health: Equitable Healthier Cities for Sustainable Development. World Health Organization. https://apps.who.int/iris/handle/10665/204715.

Özkaynak, H., Baxter, L.K., Dionisio, K.L., Burke, J., 2013. Air pollution exposure prediction approaches used in air pollution epidemiology studies. Journal of exposure science & environmental epidemiology, 23(6), 566-572. doi.org/10.1038/jes.2013.15

Pala, D., Caldarone, A.A., Franzini, M., Malovini, A., Larizza, C., Casella, V., Bellazzi, R., 2020. Deep Learning to Unveil Correlations between Urban Landscape and Population Health. Sensors, 20(7), p.2105. https://doi.org/10.3390/s20072105.

Pala, D., Pagán, J., Parimbelli, E., Rocca, M.T., Bellazzi, R., Casella, V., 2019. Spatial enablement to support environmental, demographic, socioeconomics and health data integration and analysis for big cities: A case study with asthma hospitalizations in New York City. Frontiers in medicine, 6, p.84. https://doi.org/10.3389/fmed.2019.00084

Purple Air website, https://www.purpleair.com/sensors (Accessed April 2020)

Sanchez, M., Milà, C., Sreekanth, V., Balakrishnan, K., Sambandam, S., Nieuwenhuijsen, M., Kinra, S., Marshall, J.D., Tonne, C., 2019. Personal exposure to particulate matter in peri-urban India: predictors and association with ambient concentration at residence. Journal of exposure science & environmental epidemiology, 1-10. doi.org/10.1038/s41370-019-0150-5

Wilson, A. Nickisch, H., 2015. Kernel interpolation for scalable structured Gaussian processes (KISS-GP). In International Conference on Machine Learning, 1775-1784.