Exploring the potential of citizen sensing for urban air quality monitoring on the example of Kyiv, Ukraine

D. Svidzinska *(Taras Shevchenko national university of Kyiv)*

SUMMARY

The system of urban air pollution monitoring is an important tool to observe, model, and manage air quality. The existing system of air pollution monitoring in Kyiv requires modification; at the same time, a citizen network of urban air quality monitoring has been evolving recently. As this alternative system is becoming an important source of data on air pollution, the aim of this study is to perform exploratory data-driven analysis of the citizen air quality monitoring network. The citizen system of urban air quality monitoring in Kyiv includes 133 sensors from 7 citizen science projects collecting data on 18 parameters with a high temporal frequency of 1-10 minutes 24/7. The data allows for the exploration of pollution patterns at high temporal and spatial detail. Inconsistencies in temporal dimensions, data gaps due to sensors inactivity, and absence of data quality flags require the application of advanced techniques of (geo)statistical data preprocessing and analysis. Future applied developments may rely on ground data fusion with remote observations for the purposes of remote datasets verification and downscaling.

Introduction

As a large city with a population exceeding 3 million people, a high concentration of industrial pollution sources, and a constantly growing number of cars, Kyiv is showing deterioration in air quality. According to the data from Central Geophysical Observatory named after Boris Sreznevsky (CGO), between 2009-2019 mean annual composite air quality index (CAQI) in Kyiv has increased by 2.5 points (at a rate of 0.2 points/year). If the rate persists, by 2025 mean annual CAQI may reach the unhealthy value of 11.3±2.1 (95% confidence interval).

In this context, the system of urban air pollution monitoring becomes an important tool to observe, model, manage air quality, and assess human exposure. The existing system of air pollution monitoring in Kyiv requires modification in the terms of spatial density and temporal frequency of observations, number of pollutants under control (Dyachuk et al. 2019). At the same time, a citizen network of urban air quality monitoring has been evolving in Kyiv for the last couple of years. As this alternative system is becoming an important source of data on air pollution, the specific objectives of this study are (1) to compare state and citizen air quality monitoring systems; (2) to perform exploratory citizen data analysis to assess data availability, abundance, and completeness; (3) to explore the potential of the data for temporal and spatial air pollution analysis.

Theory

The state system of air monitoring includes 16 CGO stations. The alternative system under the umbrella of the SaveEcoBot (www.saveecobot.com) initiative currently brings together 7 citizen sensing projects with 133 stations. According to Fig. 1-A, the stations are located unevenly, with the highest concentration in the central and densely-built parts of the city and along major highways, leaving outer parts of the city underrepresented.

Jointly, both monitoring networks collect data on over 20 air quality parameters (Fig. 1-B). The CGO stations register data on 20 pollutants including heavy metals (Cd, Cr, Cu, Fe, Mn, Ni, Pb, Zn). Some of the pollutants (C_6H_6O, H_2S, soot) are mentioned in the technical documentation of the CGO network but not included in Fig. 1-B because of the data unavailability. Also, not all parameters are simultaneously measured on all stations. For example, all 16 stations measure CO concentrations but only 1 station registers data on NO.

Citizen sensors collect data on 18 parameters, but, again, the number of stations for each parameter varies (Fig. 1-B). Among the most extensively registered pollutants are PM_{10} and $PM_{2.5}$ concentrations (primarily via EcoCity, lufdaten.info, and Save Dnipro networks). Other widely measured (over 20 stations) substances are CO and NO_2. So far, Kyiv Smart City, EcoCity, and Taras Shevchenko National University of Kyiv collect the most diverse sets of air quality parameters among all the projects (16, 12, and 9 respectively), but, with 30 active stations, EcoCity exceeds the other networks on the volume of data collected. Aside from pollutants, the majority of the sensors register air temperature, humidity, and atmospheric pressure. This additional data may be of use to discern interdependencies between pollutant concentrations and meteorological variables and to develop more accurate air quality prediction models.

Temporal dimensions of the data from both monitoring systems vary in the terms of frequency and span. CGO stations register concentrations 3-4 times per day (01 AM, 07 AM, 13 PM, 19 PM), 6 days per week. Over 90% of the citizen sensors are capable to automatically register data with a time step of 1-10 minutes 24/7. CGO collects data since at least the 1990s which makes it an immense source of the historical record. In the citizen network, overall data collection spans usually do not exceed 1-2 years, and data records may contain gaps of different duration (minutes, hours, days). For example, the Kyiv Smart City network started collecting the data on 31.05.2019 and stopped 06.12.2019 (some stations continued working till January 2020). The mission was expected to be resumed in June 2020, but as of September 2020, the sensors are still inactive.

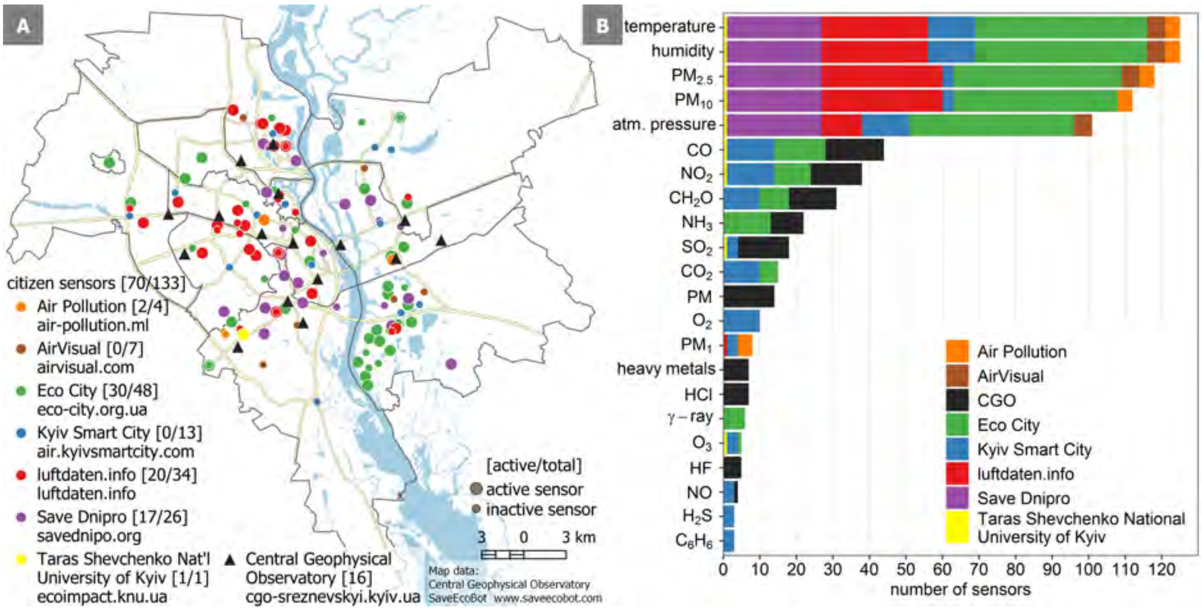

***Figure 1** Air quality monitoring stations in Kyiv (A); parameters they measure (B) as of August 2020*

In the context of data-driven decisions making, data access policies directly define the practical uptake and value of the monitoring results. Currently, state and citizen monitoring networks adopted contrast approaches to the data distribution. CGO provides its data on-demand on a commercial basis usually in aggregated and not machine-readable format. On the contrary, citizen data openly available online under Creative Commons Attribution 4.0 International Public License via saveecobot.com platform (both, direct download and API are supported) in machine-readable CSV-files. Alternatively, users may search and download data via projects' web-sites which also usually do not restrict data access.

Examples

As an example of a single sensor-based temporal analysis, I selected a sensor from the Kyiv Smart City network. The sensor is an automatic station registering data with a 5-minute time step and located at 5, Saliutna Street, Shevchenkivskyi district. The overall data record for the station contains 1,539,636 entries (over 118,000 entries per parameter) on 10 pollutants (PM_1, $PM_{2.5}$, PM_{10}, NO, NO_2, O_3, CO, SO_2, H_2S, C_6H_6), humidity, pressure, and temperature. The overall data registration time span encompasses 333 days from 19.06.2019 to 16.05.2020 with 170/163 sensor active/inactive days respectively (the longest data gap of 54 days is from 12.03.2020 to 04.05.2020).

Because of the length and distribution of data gaps, the dataset is more suitable for hourly aggregation to represent 24-hour pollutants concentration dynamics rather than daily aggregation to track seasonal and annual pollution patterns. To exclude possible low-quality observations, as the data does not contain quality flags, the time series were filtered out from outliers that exceed ±1.5IQR and then aggregated into 20-minute means for hours 00-24. For O_3, H_2S, C_6H_6 only zero concentrations were registered, thus the results include 7 pollutants (Fig. 2).

Concentrations of all pollutants demonstrate clear 24-hour dynamics pattern usually with two peaks registered AM and PM. NO concentrations distribution is unimodal with an AM maximum followed by NO_2 elevated concentrations. Joint dynamics of NO_x concentrations may be related to the fact that both pollutants are rapidly inter-converted during the day and exist in quasi-equilibrium depending on the amount of the sunlight. The bimodal concentration distributions confirm the predominance of traffic as the main source of air pollution in Kyiv.

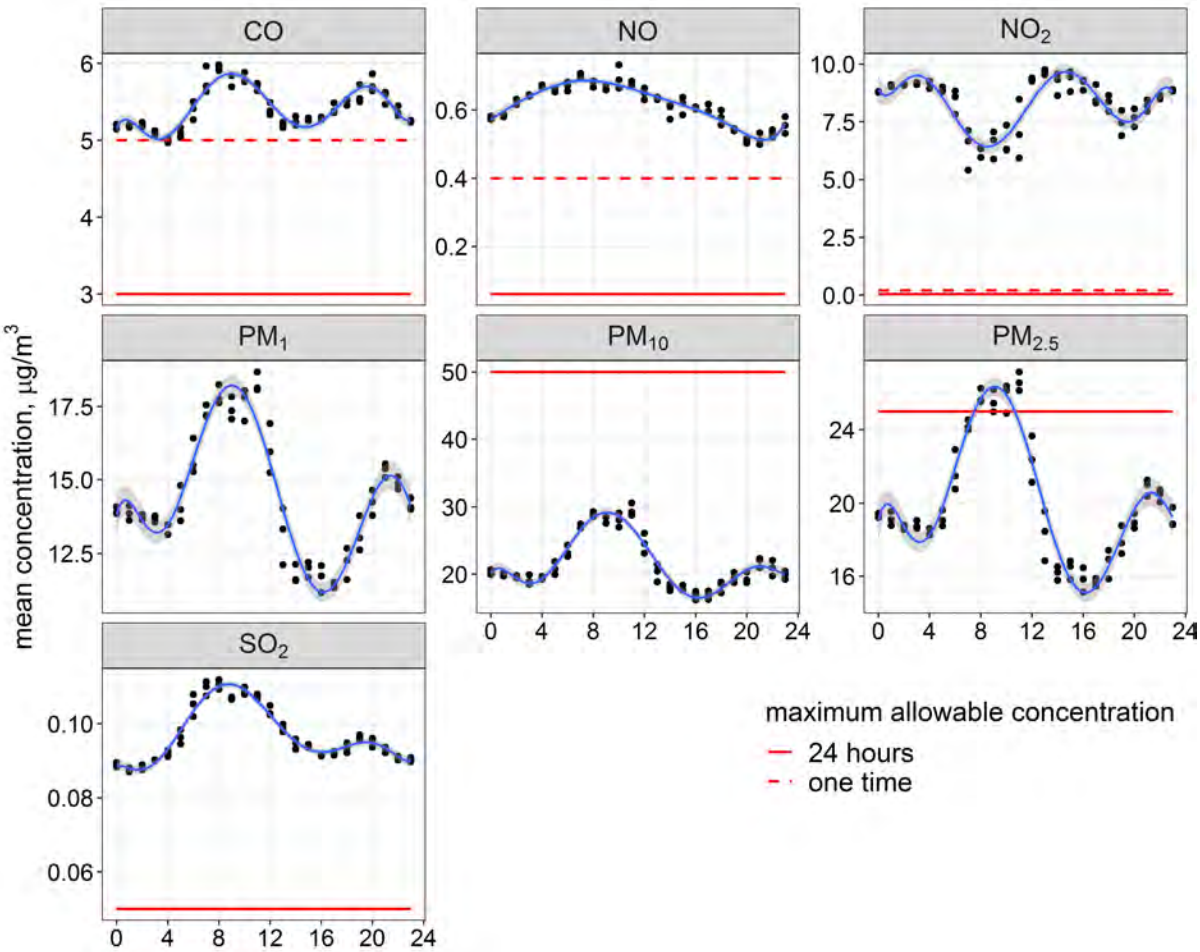

Figure 2 *24-hour dynamics of pollutants concentrations averaged from the Kyiv Smart City sensor measurements registered for 170 days from 19.06.2019 to 16.05.2020 at 5, Saliutna Street*

In general, PM concentrations do not exceed allowable limits, only $PM_{2.5}$ values during AM peak tend to be higher than 25 µg/m^3. Other pollutants significantly surpass both 24-hour and one-time maximum allowable concentrations. For SO_2 concentrations are higher than the 24-hour limit of 0.05 µg/m^3 but are lower than the one-time maximum concentration of 0.5 µg/m^3 (not shown on the plot for visual compatibility). Extremely high NO_2 concentrations registered by the sensor are a subject for further investigation.

To explore the spatial pattern of the pollutant concentration distribution, the dataset was analyzed to find the date with the maximum number of active sensors. With the total number of 80 active sensors, May 25th, 2020 was defined as the date of interest. $PM_{2.5}$ was selected as the parameter of interest because 118 sensors in total are technically capable to record its concentrations making it the most widely monitored pollutant in Kyiv. As a result, 1,233,279 data records of $PM_{2.5}$ concentrations collected by 78 sensors for the period of 00 AM-24 PM 25.05.2020 were aggregated into 24-hour means on a station-by-station basis and mapped (Fig. 3).

As of 25.05.2020 for most of the stations $PM_{2.5}$ 24-hour mean lower than the World Health Organization recommended limit of 10 μg/m^3, and no stations exceeding the EU limit value of 25 μg/m^3. The area of higher $PM_{2.5}$ concentrations lies in the south-west of the left-bank part of the city (Nizhny Sady/Lower Gardens, Osokorky). The waste incineration plant "Energy" may be responsible for such concentrations, since it's the major industrial source of air pollution located in the area. Another reason may be local microclimate and atmosphere circulation conditions formed during the day.

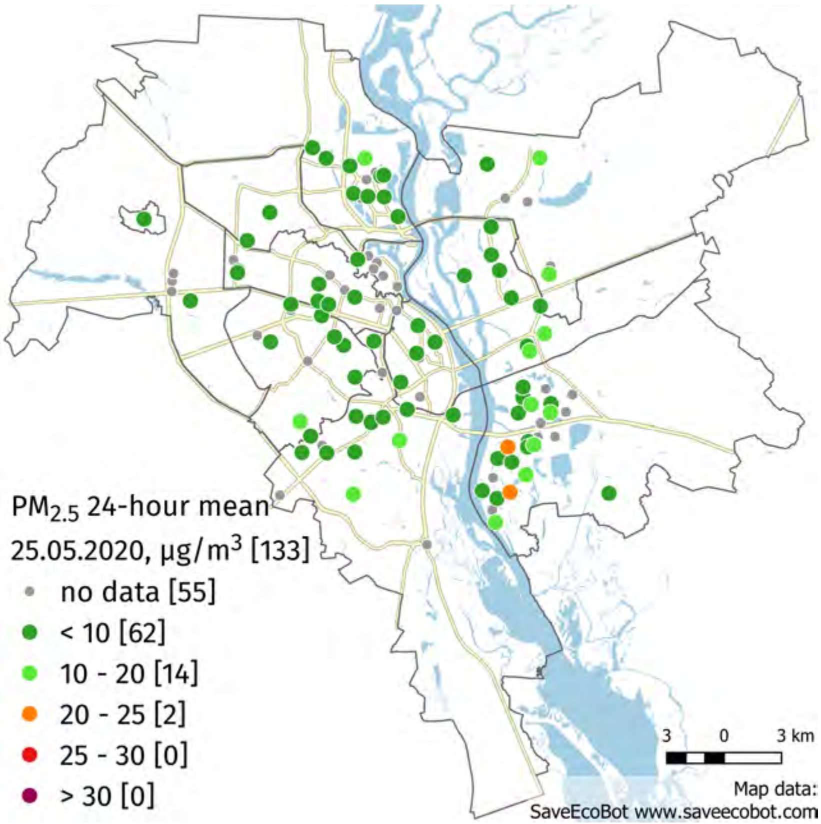

Figure 3 *Spatial distribution of $PM_{2.5}$ 24-hour mean concentrations registered for 25.05.2020*

Conclusions

The citizen system of urban air quality monitoring in Kyiv includes 133 sensors from 7 citizen science projects collecting data on 18 parameters with a high temporal frequency of 1-10 minutes 24/7. The data allows for the exploration of pollution patterns at high temporal and spatial detail. Inconsistencies in temporal dimensions, data gaps due to sensors inactivity, absence of data quality flags require the application of advanced techniques of (geo)statistical data preprocessing and analysis. Future applied developments may rely on ground data fusion with remote observations for the purposes of remote datasets verification and downscaling (Svidzinska, Korohoda 2020).

References

Dyachuk, V.A., Bashtannik, M.P., Kiptenko, E.M., Kozlenko, T.V., Nadtochii, L.M. [2019]. Study of the air pollution state monitoring and directions for its improvement in Kiev city. *Hidrolohiia, hidrokhimiia i hidroekolohiia*, **4**(55), 111-121. (in Ukrainian)

Svidzinska, D., Korohoda, N. [2020]. Study of spatiotemporal variations of summer land surface temperature in Kyiv, Ukraine using Landsat time series. *Geoinformatics: Theoretical and Applied Aspects 2020. European Association of Geoscientists & Engineers.* https://doi.org/10.3997/2214-4609.2020geo106