Инновации экоконтроля: оптимизация процесса выявления предприятий, загрязняющих воздух, с использованием данных сенсорных сетей

Аннотация.

Процесс поиска источников загрязнения воздуха по данным косвенных показателей является важным, но трудоемким процессом. В данной работе мы представляем прототип технологического решения, которое позволит инспектору по охране окружающей среды быстро оценить состояние воздуха на основе интерактивной карты качества воздуха, на которой отмечены все известные предприятия и датчики качества воздуха. Разработанное приложение также позволит инспектору делать пометки, видеть аномалии, а также прогнозировать динамику состояния с шагом в 20 минут или на сутки вперед. В приложении также можно будет увидеть динамику по дням, месяцам и годам на специальной информационной панели. Часть управления приложением реализована в виде чат-бота, который поможет инспектору придерживаться графика проверок, сообщит об аномалии, а также сделает отметку о ее геопозиции. Представленное решение обладает оригинальным набором функций, необходимых для решения задачи поиска источников загрязнения.

Ключевые слова: Flutter, Python, инновации, загрязнение воздуха

Vinogradova Elizaveta,
MSc in bioinformatics, Saint-Petersburg
National Research University of Information Technologies, Mechanics and Optics

Savelyeva Olesja Vyacheslavovna
S. P. Botkin State Clinical Hospital
Marketing Specialist
GSc in information security, Rostov State University of Economics

Carlo Julia,
JSC “Greenatom”
Senior Specialist
MSc in design, Institute of Business and Design [B&D]

Pritesh Mahendrabhai Soni
BSc in Computer Systems, Riga Technical University

Ecocontrol Innovation: Optimizing the Identification of Air Polluting Enterprises Using Sensor Network Data

Abstract.

The process of finding sources of air pollution from indirect indicator data is an important yet time-consuming process. In this paper, we present a prototype of a technological solution that enables an environmental inspector to quickly assess air quality based on an interactive air quality map on which all known enterprises and air quality sensors are marked. The developed application also allows the inspector to take notes, see anomalies and predict air quality changes in 20-minute or 24-hour increments. The app provides a view of the dynamics by day, month and year on a special dashboard. Part of the management of

the application is implemented in the form of a chatbot that will help the inspector adhere to the inspection schedule, report an anomaly and make a note of its geo-positioning. The presented solution has an original set of functions necessary to solve the problem of finding pollution sources.

Keywords: Flutter, Python, innovation, air pollution

Introduction.

There is a certain correlation between air pollution and various human diseases, especially lung diseases [Kelly, F.J. *et al.,* 2012*;* Rohde, R.A. and Muller, R.A., 2015; Apte, J.S. *et al.*, 2017; Carlsten, C. *et al.*, 2020; Shaddick, G. *et al.,* 2020; Venter; Z.S. *et al.,* 2020; Juginović, A. *et al.,* 2021; Liu, B. *et al.*, 2021], so it is important to monitor the concentration of airborne pollutants. In addition, timely detection of pollution and locating the source of it are essential for environmental protection [Carlsten, C. *et al.*, 2020; Juginović, A. *et al.,* 2021].

Therefore, recently *inexpensive sensor air quality monitoring networks* have been increasingly used to study the spatial and temporal variation of pollutant concentrations in the air [Kelly, F.J. *et al.,* 2012*;* Bellinger, C. *et al.,* 2017; van Zoest, V. M., Stein, A. and Hoek, G., 2018; Carlsten, C. *et al.*, 2020; Shaddick, G. *et al.,* 2020].

One way to describe such air quality monitoring networks is to describe their necessary characteristics, as was done in [Li, X. et al., 2021]. Li, X. et al. argue that an air quality monitoring network should be *representative*, *comparable*, and *stable*. It should reflect actual air quality levels, provide consistent data, provide information across the territory and all relevant objects, be stable in time and space, and allow data analysis and prediction. However, due to the costs associated with the construction, operation, and maintenance of air quality sensors [Kelly, F.J. *et al.,* 2012*;* Luo, X. and Yang, J., 2019], the actual temporal and spatial accuracy of *general monitoring is limited*, although the high accuracy of monitoring equipment makes it possible to establish representative and stable air quality monitoring networks. This makes it particularly difficult to trace the true sources of air pollution.

Problem description.

Urbanization, more intensive energy consumption and increased emissions from transport and industrial sources are constantly deteriorating air quality. As a result, people in both developed and developing countries [Juginović, A. *et al.,* 2021] are exposed to a greater variety of air pollutants, many of which are in unhealthy concentrations in many urban areas.

The health response to increased outdoor air pollution varies among individuals and population subgroups in a way that individual susceptibility affects the observed level of health effects and the rate of symptom exacerbation with increasing air pollution concentrations [Kelly, F.J. et al. 2012; Juginović, A. *et al.,* 2021]. Reasons why some individuals appear to be more susceptible may include genetic predisposition, chronic respiratory and cardiovascular disease, metabolic disorders such as diabetes, or suboptimal levels of dietary antioxidants [Kelly, F.J. et al., 2012]. Age also affects response, as children have developing lungs and immune systems, while older adults tend to face accumulation of chronic diseases and aging of body systems [Kelly, F.J. et al., 2012]. Other factors that may contribute to greater susceptibility include increased physical activity and thus increased ventilation, and social deprivation, which may be due to factors such as higher rates of chronic disease, poor nutrition, and greater exposure [Kelly, F.J. et al., 2012; Juginović, A. *et al.,* 2021]

Air pollution monitoring networks around the world report local ambient air pollution levels in real time and can provide information on when and where air pollutant levels exceed levels that are considered to pose health risks [Carlsten, C. *et al.*, 2020].

However, one cannot disagree that air monitoring networks are highly dependent on the number of sensors, the accuracy of measurements and the reliability of the equipment. As shown in [Luo, X. and Yang, J., 2019], there are problems in monitoring networks related to both the number of sensors (nodes) in the network and the number of sensors measuring different air quality indicators. In addition, there are problems with the efficiency of the power supply and communication infrastructure. However, the biggest challenge for existing air monitoring networks is the problem of accurately locating the sources of air pollution and the problem of quickly alerting and coordinating the work of air quality inspectors. The latter is important due to the instability of weather conditions [Kelly, F.J. *et al.,* 2012*;* Lewis, T. and Bhaganagar, K., 2021; Liu, B. et al. 2021] and the irregularity of pollutant source activity, intertwined with the likely malevolent intentions of violators who do not want to be identified.

Research background.

Air quality monitoring is critical for environmental protection [Fowler, D. *et al.,* 2020; Carlsten, C. *et al.*, 2020].

Air quality monitoring is performed using *data from a network of sensors* with a help of specially developed *monitoring systems*. Although these terms are sometimes used interchangeably.

For any air quality monitoring system, in order to characterize releases, information must be collected through a process called source term evaluation (STE), which aims to estimate several parameters, including source location, release intensity, release duration, pollutant type, and dispersion [Lewis, T. and Bhaganagar, K., 2021]. Continuous monitoring in the STE process helps to address three main issues, including monitoring the state of the environment, detecting anomalies (releases and errors), and locating the source of contamination [Luo, X. and Yang, J., 2019].

During the monitoring process, data is collected and transferred to the data storage. The task of anomaly detection is then to determine the fact of contamination by analyzing the data. The last task, localization, is to determine the location of the pollution source.

In the following sections, these main tasks are described in more detail.

Monitoring.

Environmental monitoring is essentially the process of collecting and analyzing data.

Monitoring systems are mostly hierarchical. The first level is the monitoring networks. Sensor nodes collect information about the environment and transmit data to the receiving node. The second level is the data transmitting level. The receiving node uploads data to the data storage via wired or wireless networks [Luo, X. and Yang, J., 2019]. The main differences between the different monitoring systems offered are in nodes and networks, such as different sensor platforms, different network topologies, different communication protocols, wired or wireless connection [Luo, X. and Yang, J., 2019].

Additionally, contaminants are not always on the surface of the water. *Different* sensor networks are used to monitor underwater pollution. These systems usually have a receiver node floating on the water surface that collects information from underwater sensor nodes and transmits the data to a processing center [Luo, X. and Yang, J., 2019].

Anomaly detection.

In some monitoring tasks, such as environmental monitoring and pollution detection, it is often necessary to know exactly whether contamination is present or not. And to solve this problem, methods for detecting anomalies have been developed.

Anomaly detection (outlier detection) is defined as the detection of values that are statistically significantly different from the expected value at a given time and place [Luo, X. and Yang, J., 2019]. Anomaly detection is important not only for detecting air pollution, but also for correcting errors related to, for example, extreme weather conditions of low wind speed and high atmospheric stability [van Zoest, V.M., Stein, A. and Hoek, G., 2018].

One of the most common methods for detecting anomalies in air quality data is the simple threshold method [Kelly, F.J. *et al.,* 2012*;* Luo, X. and Yang, J., 2019; Carlsten, C. *et al.*, 2020]. Although the threshold method is easy to implement and interpret, it does not provide the necessary flexibility; therefore, other methods have been developed, mainly to improve the basic characteristics of the threshold method. The methods for detecting anomalies in air quality data are described in detail below.

There are many approaches to detecting anomalies. To detect an anomaly, one can compare observations with their temporal neighborhoods. This method is suitable for measurements with minimal variability. For observations with high variability and/or low resolution, one can use a method that compares curves of various parameters, for example, comparing large and small particles at the same time. Another method that is resistant to temporal changes in data and used to detect anomalies is a method based on taking into account spatial characteristics. When using this outlier detection method, an observation is compared with observations in its spatial and temporal neighborhood. However, this method only works well for observations with low spatial variability [van Zoest, V.M., Stein, A. and Hoek, G., 2018].

Therefore, the most suitable method for detecting anomalies in urban conditions is the hybrid approach, which takes into account both spatial and temporal variability of air quality data. One such approach is described in a work [van Zoest, V.M., Stein, A. and Hoek, G., 2018], where researchers used nitrogen dioxide data from a network of air quality sensors located in the city of Eindhoven, the Netherlands.

The proposed method for detecting outliers [van Zoest, V.M., Stein, A. and Hoek, G., 2018] is based on checking whether the observed concentration values fall within a given confidence interval (95% or 99.7%) for the grouped by temporal, spatial and spatio-temporal characteristics data. In their work, the researchers implied that the observations should be independent and normally distributed. They showed that in urban conditions, readings grouped temporally were strongly dependent on the time of day (which is a valid observation, and also mentioned by Kelly, F.J. *et al.*) and readings grouped spatially were strongly dependent on permanent sources of pollution, such as highways. To overcome these limitations that make it impossible to reliably detect anomalies, van Zoest, V.M., Stein, A. and Hoek, G. further developed a special method to separate data by temporal and spatial characteristics. In total, they describe 16 spatio-temporal categories to separate data for different levels of air pollution: locations of increased urban traffic, urban background locations, presence of highways, distance from the center, a day divided into four intervals, locations where sensors are close to specific types of land use, and weekdays and weekends are also separated. For each spatio-temporal category, three steps were taken to establish anomaly thresholds: (1) transform the measurements using square root extraction to make the distribution of the indicators closer to normal; (2) calculate adapted values for center and spread statistics; and (3) apply the previously described anomaly detection method, using confidence intervals adapted to the new parameters. The computed thresholds are subsequently not used for the actual anomaly detection, but as an approximation of the thresholds for each spatio-temporal category. As a result of the analysis, van Zoest, V. M., Stein, A. and Hoek, G. determined that the proportion of outliers was approximately the same for all categories, without large deviations. It should be remembered here that due to the spatial classification introduced by [van Zoest, V. M., Stein, A. and Hoek, G., 2018], some concentration values are considered as outliers in urban background areas, while they are not outliers in urban traffic areas. The robust approach proposed by van Zoest, V. M., Stein, A. and Hoek, G. is one of the promising methods to deal with air quality data, taking into account the micro and macro variability of the data by separating it using spatio-temporal characteristics.

Another useful perspective on anomaly detection in sensory data is described in the paper by Luo, X. and Yang, J. The paper [Luo, X. and Yang, J., 2019] describes anomaly detection in terms of change detection methods. In statistical analysis, change detection methods attempt to identify moments when the probability distribution of a random process or time series changes. In general, the problem involves both detecting whether a change has occurred, or whether more than one change may have occurred, and determining the timing of any such change.

Although anomaly detection methods are an important part of the air quality monitoring pipeline, due to the nature of the observations being strongly influenced by the low density of sensors in most regions, it is not the anomaly detection methods but the source localization methods that have received the most attention in most works. The following sections describe the two main groups of source localization methods in more detail.

Source localization.

The models used to determine the sources of pollutants are broadly divided into two groups. The first group consists of physical models (dispersion models) that use atmospheric system theory to simulate the physical and chemical processes of pollutants in a given region. The other group (nonphysical or receptor models) approaches include statistical models and machine learning methods, iterative algorithms, and heuristics-based algorithms [Kelly, F.J. *et al.,* 2012*;* Luo, X. and Yang, J., 2019; Liu, B. *et al.*, 2021].

The flow field describes the direction and size of the flow at any point in the substance. Contaminants migrate through the flow field, so most works try to locate the source of contamination based on the principle of contamination dispersion [Luo, X. and Yang, J., 2019]. These approaches fall into the group of physical models.

When using these physical models, diffusion models are used to localize the source of contamination. Pollution sources in such models are considered as point sources, and the transfer of substances from the source of diffusion is described by diffusion equations. As a rule, the diffusion coefficient in air is considered to be greater than in water [Luo, X. and Yang, J., 2019].

In non-physical models various approaches are used.

In the maximum monitoring value point approach (MPA), the sensor node with the maximum control value is considered to be very close to the source of contamination.In this approach, the location of the sensor node in the network with the largest control value is considered the source of contamination [Luo, X. and Yang, J., 2019].

In the approach using the earliest detection point (EPA), the position of the source is defined as the location of the sensor node that first detects contamination [Luo, X. and Yang, J., 2019].

The MPA and EPA methods refer to the CPA (approximation to the nearest point) methods.

In the centroid localization algorithm (centroid algorithm): all sensors that have detected contamination are clustered and the centroid of the area in which they are located is considered as the location of the source [Luo, X. and Yang, J., 2019].

In both cases, the physical and non-physical algorithms are limited either by the resolution of the sensor network or by the available precision of the diffusion models.

For localization algorithms based on network observations rather than contaminant immigration laws, the accuracy depends on the number and quality of sensor nodes in the network. And at the same time, since all diffusion models rely on ideal assumptions, they do not represent the diffusion process in the real environment very accurately. As a consequence, the localization based on diffusion models also has limitations.

Therefore, without analyzing the conditions under which these algorithms are used, it is difficult to say which localization algorithm works better.

Problems of localizing pollution sources.

As mentioned above, the use of sensor networks for air quality monitoring presents many challenges. The main problems are related to the price of the equipment and its maintenance [Luo, X. and Yang, J., 2019; Liu, B. *et al.,* 2021], as well as the complexity of the environment in which measurements take place.

A more detailed description of the problem list is given in a review [Luo, X. and Yang, J., 2019]. According to Luo, X. and Yang, J., another important characteristic of pollution sources is the possibility of stochastic events. In some cases, there is no regularity in the migration of pollutants and the characteristics of the source vary randomly. In these cases, the use of sensor values to detect pollution sources is a problem. Furthermore, Luo, X. and Yang, J. added the complexity of physical models of pollution propagation to the list of problems. According to the authors of the review, this problem is firstly related to the nonlinearity of the pollutant source localization problem (for sensor indicators), which introduces many local extrema in the solution domain. Secondly, the complexity of the physical models used to describe the contaminant migration makes it necessary to use simplified properties of the environment. Therefore, in order to simplify the research task, most of the work on pollution monitoring using sensor networks ignores advection and chemical reactions, as well as transport and fate models of pollutants.

Forecasting.

An additional task that is currently required of any air quality monitoring systems [Lewis, T. and Bhaganagar, K., 2021; Liu, B. *et al.,* 2021] is the task of air quality prediction. Forecasting methods in such systems are mainly used to predict the dynamics of air quality changes [Liu, B. *et al.,* 2021] after an anomaly has been confirmed.

Novelty of the research.

In this study, we illustrated how a robotic assistant can be used to improve the performance of a semi-automatic contaminant detection system.

Research hypothesis.

In this study, we address the problem of developing air quality monitoring systems, the importance of which has been comprehensively described in many recent works [see survey by Luo, X. and Yang, J.]. In the present work, however, we hypothesize that the use of additional tools to control and interact with the monitoring system, such as a robotic assistant, can improve the performance of such semi-automated pollution source detection systems.

Aim of research.

The goal of this study was to develop a prototype of a monitoring and alert system suitable for semi-automated operation, with active use of data from the city and regional sensor network and data from open sources, optimized to facilitate and enhance the work of environmental inspectors by using additional interaction tools, in particular a robotic assistant.

Research objectives.

In this study, when developing a prototype of a monitoring system, we focused our attention on the following five tasks:

(1) Development of an algorithm for localization of pollution sources based on algorithms (MPA, EPA)
(2) Development of an entity matching algorithm for compiling lists of possible violators according to open sources
(3) Application of geocoding based on discrete grids, such as Uber H3
(4) Improved interaction with data. Data visualizations on the interactive map and dashboards
(5) Development of an assistant robot for an environmental inspector (chatbot development). An additional tool for managing and interacting with the developed prototype of the monitoring system

Methods.

To solve these tasks, we used modern methods of developing cross-platform applications.

We use the Flutter library [Google Inc., 2017] for web and mobile development.

Flutter is Google's portable UI toolkit for making natively compiled applications for mobile, web, and desktop from a single code base. Flutter can be used with existing code, is used by developers and organizations around the world, and is free and open source [Google Inc., 2017].

We used the TensorFlow machine learning framework [Abadi, M. *et al.*, 2015] to develop our machine learning algorithms, which are mentioned results section below.

TensorFlow is a machine learning framework that can run in large-scale and heterogeneous environments.TensorFlow uses data flow graphs to represent computations, shared state, and operations that change that state.TensorFlow supports a variety of applications with a focus on deep neural network training and inference. Several Google services use TensorFlow in production and it has been widely used in machine learning research [Abadi, M. *et al.*, 2015].

Finally, we use the Docker [Merkel, D., 2014] platform for virtualization to deliver our software.

Docker is a software for automating the deployment and management of applications in environments that support containerization, and is an application containerizer. It allows to package an application with its entire environment [Merkel, D., 2014].

Data.

In this project, we used data from air control stations in the Chelyabinsk region in Russia, as well as open data on enterprises in this region.

Results. Anomaly detection.

In this study, we investigated anomaly detection methods for sensor network data used to monitor air quality based on reviews by Luo, X. and Yang, J. and Liu, B., and promising work by van Zoest, V. M., Stein, A., and Hoek, G., in which a hybrid method for detecting anomalies is described. In the work of van Zoest, V. M., Stein, A., and Hoek, G., an intuitive, but also powerful, algorithm for detecting anomalies in sensor measurements is proposed that takes into account both temporal and spatial features. Since we believe in the correctness of the statements and conclusions presented by van Zoest, V. M., Stein, A. and Hoek, G., it appeared to be correct to implement a new hybrid approach for anomaly detection in our application. However, unfortunately for our data, this approach did not bring a significant improvement in the quality of anomaly detection, and even on the contrary, many events were no longer detected. We attribute this to the fact that for our data, there are not enough sensors while they cover a large area.

Therefore, in the current work, we used the threshold method to detect anomalies.

Source localization.

For our monitoring system, we decided to implement one of the nearest point approximation methods, in particular, the earliest detection point algorithm, which we use in conjunction with a specially developed entity matching algorithm.

Entity matching algorithm.

The main problem in accurately searching for businesses that are sources of air pollution, besides the low territorial coverage with various air quality sensors, lies in the fact that the location of pollution does not always correspond to the official address of the business, as well as in the irregular naming in existing databases.

In this work, we developed a special hierarchical entity matching algorithm that uses data from open sources to compile a list of the most likely sources of air pollution.

The developed hierarchical entity matching algorithm works in two stages: first, it finds the possible objects in the polluted area and then, given the proximity of the objects that could be sources of pollution (e.g., businesses), it performs a ranking. In the first step, the algorithm finds possible entities from public databases using taxpayer identification numbers, company names and owners, product or service characteristics, and other data that are mostly in text form.

It is in the first step of the algorithm that the mapping of objects occurs. We use three methods to match records and information about objects and map them to an area. First, the algorithm uses Hamming distance with basic features such as company name and owner name, then it uses SentenceTransformer [Reimers, N. and Gurevych, I., 2020] embedding-based matching, finally, it uses another type of embedding-based matching, empowered by an autoencoder that we previously developed and trained on existing data. In the latter case, matching is performed on the latent representations of records obtained from the encoding head of the autoencoder.

In the second stage, the algorithm ranks the found objects according to their proximity to the possible source of contamination.

The developed method proved to be quite effective on the available data.

Geocoding and visualization.

To visualize the data, we preliminarily mapped more than 100 thousand addresses of enterprises with their coordinates, and then with the mapping on a Uber H3 discrete grid, which is described in work [Brodsky, I., 2018]. After displaying the coordinates of the enterprises, we built an interactive heat map of the studied territory, which is further used in the monitoring system. The constructed interactive map based on the Uber H3 technology has a number of advantages over traditional clustering methods, although it requires regular updating.

Interactive map.

Further, based on official data from air pollution control stations, using interpolation, we constructed a map of air pollution at each data point. For interpolation, we used piecewise cubic interpolation (see Fig. 1-3) from the SciPy library [Virtanen, P. *et al.*, 2020].

We have also implemented analysis tools for all types of pollution and the ability to dynamically add new stations.

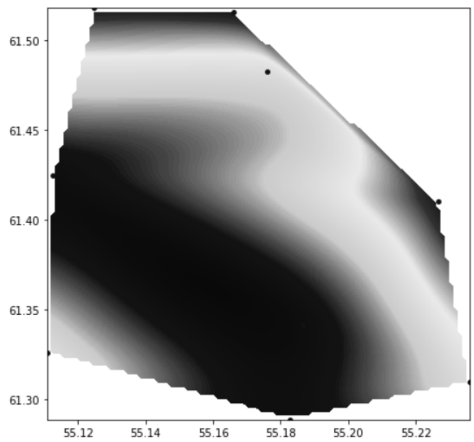

Fig.1. The result of piecewise cubic interpolation

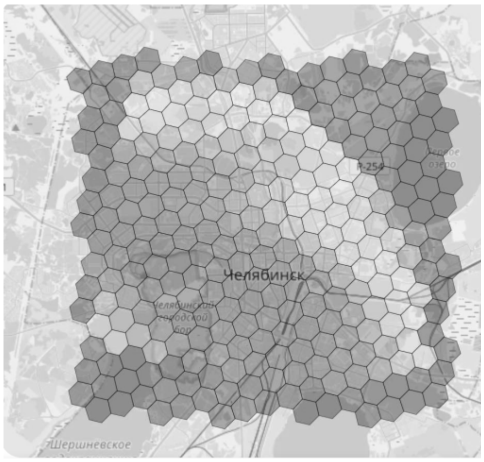

Fig.2. The result of combining Uber H3 grid technology and piecewise cubic interpolation results [the same area as in Fig.1]

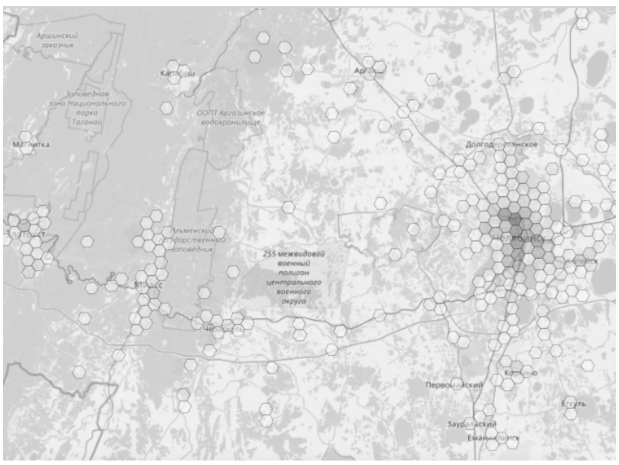

Fig.3. The result of combining Uber H3 grid technology and piecewise cubic interpolation results [entire territory]

Dashboard and robotic assistant.

We have also developed an information and analytical dashboard where one can see all the necessary information about organizations and enterprises (see Fig. 4-5) and a chatbot assistant (see Fig. 6). With the help of the bot, environmental inspectors can quickly respond to events and make the necessary decisions. The bot in the mobile app allows field employees to send their location to the system (see Fig. 6) and view the history of site visits by other employees and fill out the necessary report.

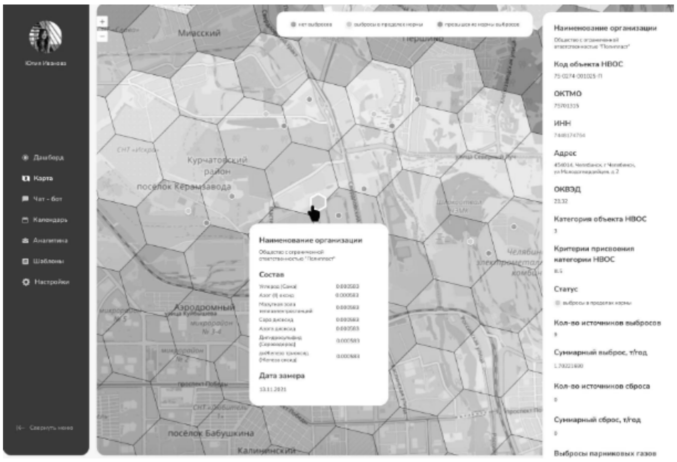

Fig.4. Interaction with the interactive map, which provides information about the requested object

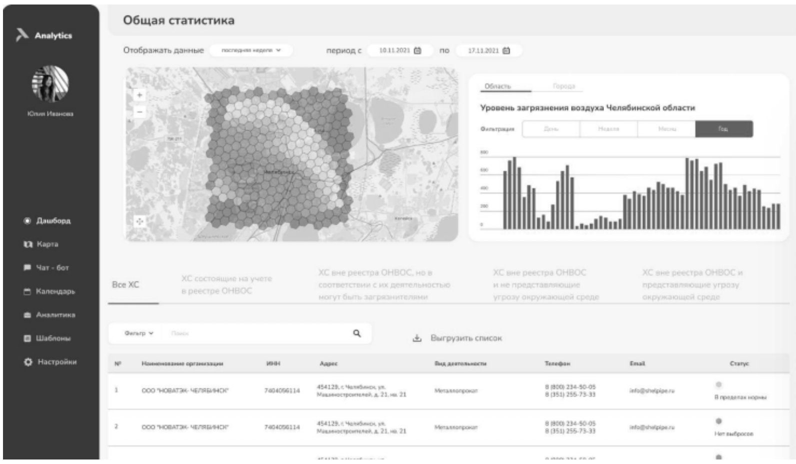

Fig.5. Main view of the developed system and dashboards

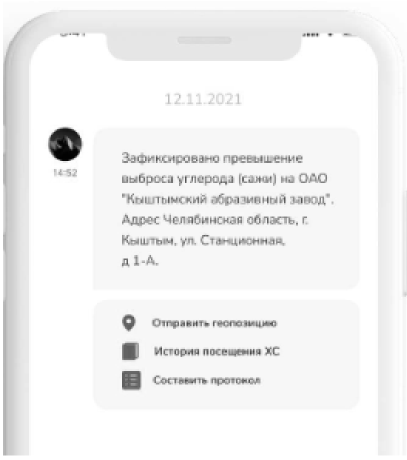

Fig.6. View of scenario interaction with the robotic assistant

Discussion.

Inexpensive air quality sensors can help one to monitor the concentration of pollutants in the air in real time [Rohde, R.A. and Muller, R.A., 2015; Shaddick, G. *et al.,* 2020].

In this work, we have shown that if the monitoring area is large and the location of the pollutant source is unknown, then the question of where to place the sensor nodes must be properly considered in practical applications. We agree with Luo, X. and Yang, J. that this is because the number of sensors in any area is distributed in a poisson manner. Therefore, we can say that a node detects contamination in time only when it is close to the source. Therefore, the performance of the detection method is always related to the density of nodes. Although, high accuracy detection methods with sparse node density is a problem that needs further research [Kelly, F.J. *et al.,* 2012*;* Luo, X. and Yang, J., 2019].

The results show that anomaly detection considering temporal and spatial characteristics [van Zoest, V. M., Stein, A. and Hoek, G., 2018] as well as expected levels is one of the most reliable methods for detecting emissions in areas with highly variable spatial and temporal air pollutant concentrations.

As a result of our work, we have developed a prototype of the monitoring system and added an additional tool to interact with the system - a chatbot assistant. Currently, we hope to confirm through more thorough testing our hypothesis that additional tools for interacting with the monitoring system can improve the quality of environmental inspectors' work.

Conclusion.

Pollution monitoring is important for protecting the environment and public health [Shaddick, G. *et al.,* 2020; Venter; Juginović, A. *et al.,* 2021]. Recently, there has been an increased interest in the topic of air pollution, which has led to certain public and political action to begin successfully reducing air pollution levels [Shaddick, G. *et al.,* 2020; Venter; Juginović, A. *et al.,* 2021]. However, even in countries with the cleanest air, large numbers of people are exposed to harmful levels of air pollution [Shaddick, G. *et al.,* 2020].

Whether used directly by environmental inspectors or not, air quality monitoring systems can increase public awareness of air pollution levels and encourage protective behaviors during periods of high air pollution, especially among at-risk individuals.

Although poor air quality can have a significant impact on human health [Kelly, F.J. *et al.,* 2012*;* Rohde, R.A. and Muller, R.A., 2015; Apte, J.S. *et al.*, 2017; Carlsten, C. *et al.*, 2020; Shaddick, G. *et al.,* 2020; Venter; Venter, Z.S. *et al.,* 2020; Liu, B. *et al.*, 2021], studies have shown that the public lacks awareness of the link between air pollution and poor health, as well as a lack of knowledge about air quality information. Real-time interactive maps can provide short-term health risk assessments based on current pollution levels in order to limit prolonged or heavy outdoor activity during times of high air pollution levels.

Air pollution indicators can be used to generate other longer-term risk interactive maps to estimate chronic air pollution exposures for patients at home and at work, and can help patients avoid high air pollution routes [Kelly, F.J. *et al., 2012;* Carlsten, C. *et al.,* 2020].

At the same time, in addition to interactive maps, other interactive tools can be used to engage users and experts through timely notifications. We believe that one such tool could be chatbots that can notify users with mobile devices.

In this work, we developed a prototype air quality monitoring and notification system suitable for semi-automated operation that actively uses data from municipal and regional sensor networks and data from open sources to facilitate and improve the quality of environmental inspectors' work. Although the system we have developed is designed for professionals, its less functional versions can still be useful to ordinary users.

Data availability.

The project page can be found https://github.com/OlesyaJ/super-data-eco**.**

References.

1. Abadi, M. *et al.* (2015) 'TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems'. Available at: https://www.tensorflow.org/.

2. Apte, J.S. *et al.* (2017) 'High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data', *Environmental Science & Technology*, 51(12), pp. 6999–7008. doi:10.1021/acs.est.7b00891.

3. Bellinger, C. *et al.* (2017) 'A systematic review of data mining and machine learning for air pollution epidemiology', *BMC Public Health*, 17(1), p. 907. doi:10.1186/s12889-017-4914-3.

4. Brodsky, I. (2018) *H3: Uber's Hexagonal Hierarchical Spatial Index*. Available at: https://eng.uber.com/h3/.

5. Carlsten, C. *et al.* (2020) 'Personal strategies to minimise effects of air pollution on respiratory health: advice for providers, patients and the public', *European Respiratory Journal*, 55(6). doi:10.1183/13993003.02056-2019.

6. Fowler, D. *et al.* (2020) 'A chronology of global air quality', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378(2183), p. 20190314. doi:10.1098/rsta.2019.0314.

7. *Flutter is Google's SDK for crafting beautiful, fast user experiences for mobile, web, and desktop from a single codebase. (2017). Google Inc. Available at: https://api.flutter.dev/.*

8. *Juginović, A. et al. (2021) 'Health impacts of air pollution exposure from 1990 to 2019 in 43 European countries', Scientific Reports, 11(1), p. 22516. doi:10.1038/s41598-021-01802-5.*

9. *Kelly, F.J. et al. (2012) 'Monitoring air pollution: Use of early warning systems for public health', Respirology, 17(1), pp. 7–19. doi:10.1111/j.1440-1843.2011.02065.x.*

10. Lewis, T. and Bhaganagar, K. (2021) 'A comprehensive review of plume source detection using unmanned vehicles for environmental sensing', *Science of The Total Environment*, 762, p. 144029. doi:10.1016/j.scitotenv.2020.144029.

11. Liu, B. *et al.* (2021) 'A data calibration method for micro air quality detectors based on a LASSO regression and NARX neural network combined model', *Scientific Reports*, 11(1), p. 21173. doi:10.1038/s41598-021-00804-7.

12. Li, X. *et al.* (2021) 'Using Sensor Network for Tracing and Locating Air Pollution Sources', *IEEE Sensors Journal*, 21(10), pp. 12162–12170. doi:10.1109/JSEN.2021.3063815.

13. Luo, X. and Yang, J. (2019) 'A Survey on Pollution Monitoring Using Sensor Networks in Environment Protection', *Journal of Sensors*. Edited by F. Meng, 2019, p. 6271206. doi:10.1155/2019/6271206.

14. van Zoest, V.M., Stein, A. and Hoek, G. (2018) 'Outlier Detection in Urban Air Quality Sensor Networks', Water, Air, & Soil Pollution, 229(4), p. 111. doi:10.1007/s11270-018-3756-7.

15. Merkel, D. (2014) 'Docker: lightweight linux containers for consistent development and deployment', *Linux journal*, 2014(239), p. 2.

16. Reimers, N. and Gurevych, I. (2020) 'Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation'.

17. Rohde, R.A. and Muller, R.A. (2015) 'Air Pollution in China: Mapping of Concentrations and Sources', *PLOS ONE*, 10(8), p. e0135749. doi:10.1371/journal.pone.0135749.

18. Shaddick, G. *et al.* (2020) 'Half the world's population are exposed to increasing air pollution', *npj Climate and Atmospheric Science*, 3(1), p. 23. doi:10.1038/s41612-020-0124-2.

19. Venter, Z.S. *et al.* (2020) 'COVID-19 lockdowns cause global air pollution declines', *Proceedings of the National Academy of Sciences*, 117(32), pp. 18984–18990. doi:10.1073/pnas.2006853117.

20. Virtanen, P. *et al.* (2020) 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python', *Nature Methods*, 17, pp. 261–272. doi:10.1038/s41592-019-0686-2.