Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 13, MARCH 2023

DOI: 10.37943/13QOJG5081

Dariya Bissengaliyeva
MSc in Information Technology, Teacher of the Department of Computer Engineering
dariya.bissengaliyeva@astanait.edu.kz, orcid.org/0000-0002-7985-162X
Astana IT University, Astana, Kazakhstan

Temirlan Amanzholov
MSc Student in Applied Data Analytics
7211127@astanait.edu.kz, orcid.org/0000-0001-5169-4346
Astana IT University, Astana, Kazakhstan

Adina Rakhimkul
MSc Student in Applied Data Analytics
7211031@astanait.edu.kz, orcid.org/0000-0002-3312-5390
Astana IT University, Astana, Kazakhstan

Didar Yedilkhan
PhD, Associate Professor of the Department of Computer Engineering,
d.yedilkhan@astanait.edu.kz, orcid.org/0000-0002-6343-5277
Astana IT University, Astana, Kazakhstan

DETERMINATION OF THE RELIABILITY OF AIR POLLUTION MEASUREMENT DATA BASED ON VEHICULAR EMISSION RECOGNIZED AS CONCOMITANT IN ASTANA

Abstract:

Air pollution is a primary global concern due to its adverse effects on human health and the environment. Accurate air pollution measurement is crucial for developing effective control strategies and evaluating their impact. In Astana, the capital of Kazakhstan, vehicular emissions are recognized as one of the significant contributors to air pollution. This study aims to determine the reliability of air pollution measurement data by examining the impact of vehicular emissions on air quality to establish effective air pollution measurement methods in the city. The study uses a combination of monitoring and modeling techniques to quantify the contribution of vehicular emissions to air pollution. The monitoring component involves the deployment of air quality monitoring stations throughout the city to measure levels of pollutants such as particulate matter (PM). The modeling component uses air dispersion models to simulate the dispersal of pollutants from vehicular emissions and predict their concentration levels in different parts of the city. This research provides insights into the effectiveness of existing air pollution control strategies and contributes to future efforts to improve air quality in Astana. Based on the collected data for a certain period, a comparative table describes the difference between the actual data from the collection points of air emission indicators. The article will likely interest researchers and policymakers concerned with air pollution and its effects on human health and the environment.

Keywords: Air Pollution, Air Quality, Emission, Data Analytics, Air Monitoring.

Introduction

Air pollution is a global environmental issue that significantly threatens human health and the environment. Various sources, including industrial activities, transportation, and

DOI: 10.37943/13QOJG5081

agricultural practices, cause it. Air pollution's effects can be severe, including respiratory problems, cardiovascular disease, and premature death. The problem of protecting and restoring the environment is currently the main one of the essential tasks of science, the development of which is stimulated in all countries of the world [1]. Accurate measurement of air pollution is crucial for understanding the extent of the problem and developing effective control strategies. It helps decision-makers to identify sources of air pollution, assess their impact on human health and the environment, and determine the effectiveness of existing control strategies.

Kazakhstan's ranking as the 23rd most polluted country in the world, with an average annual PM2.5 concentration of more than 50 μg/m3 in 2021, highlights the country's significant air pollution challenges [3]. The high levels of PM2.5 in the air are primarily due to emissions from various sources, including industrial processes, transportation, and the combustion of solid fuels for heating and cooking purposes. The country's reliance on coal-fired power plants and other facilities for electricity generation also contributes significantly to its air pollution levels. Many households and businesses rely on coal-fired boilers and stoves as they lack access to gas heating, resulting in high PM2.5 emissions [2]. As heating demand increases during the winter, coal consumption can double at coal-fired power stations and other facilities, leading to rising air pollution levels, particularly in cities and towns near these emission sources. The biggest Kazakhstani cities as, Almaty and Astana, often rank among the most polluted megacities globally, with PM2.5 levels ranging from 100 to 200 μg/m3 [4].

The World Health Organization (WHO) has set a guideline for the average daily limit of PM2.5 concentration at 15 μg/m3. This guideline is based on the latest scientific evidence about the health impacts of air pollution and is designed to protect public health [5]. It is important to note that this guideline is not a legal requirement, and many countries worldwide still have much higher levels of PM2.5 pollution than the WHO recommends. However, efforts are being made by governments, businesses, and individuals to reduce air pollution levels and bring them in line with WHO guidelines. Reducing air pollution is a complex issue that requires a multi-faceted approach, including reducing emissions from transportation and industry, promoting clean energy sources, and improving public awareness about the health impacts of air pollution [5]. Governments and individuals need to take action to reduce air pollution and improve the quality of life for those living in these areas.

Rapid economic development, exploitation of natural resources, lax environmental regulations, and weak enforcement has led to environmental degradation in many locations in Kazakhstan [6]. Astana city has experienced constant population and economic growth for several years, which inevitably led and may continue to lead to increased transport activities, urbanization, and energy demand [7].

The impact of various environmental factors on the contamination levels of urban air has been investigated in the study [8], highlighting the significance of climatic and wind conditions and the city's layout. The study reports that temperature inversions and stagnant air masses can significantly exacerbate pollution levels in urban areas, while windy conditions can help disperse pollutants and improve air quality. Furthermore, the city's layout can also contribute to high levels of air pollution. Specifically, urban canyons - areas where buildings are densely packed together - can trap pollution and hinder its dispersion, leading to a build-up of contaminants in the local atmosphere. The findings of the study [8] highlight the need for a comprehensive approach to addressing air pollution in urban areas, considering not only the sources of pollution but also the various environmental factors that can affect its concentration levels. This can include measures to reduce emissions from transportation and industrial activities, promote cleaner energy sources, and improve the urban layout to facilitate the dispersion of pollutants. Implementing such measures can improve the air quality in urban areas and reduce the negative health impacts associated with air pollution.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 13, MARCH 2023

A study [9] has identified road traffic and point sources, including various types of industries, such as thermoelectric power stations, as the primary sources of air pollution emissions in urban areas. The study highlights the significant role of gasoline and diesel-powered vehicles in emitting pollutants such as nitrogen oxides (NOx), particulate matter (PM), and volatile organic compounds (VOCs), which contribute to the deterioration of air quality. In addition, industrial activities such as power generation, manufacturing, and mining are also significant emissions sources, with power plants being a primary source of pollutants such as sulfur dioxide (SO2) and nitrogen oxides (NOx).

Recent studies have examined the trends in industrial emissions and air pollution levels in industrial cities across Kazakhstan. The study [9] analyzed data from the permitting documents of twenty-one power plants and nine metallurgical enterprises in the country and found that although industrial emissions in Kazakhstan are declining, air pollution levels in cities such as Temirtau and Ekibastuz remain high. The study has emphasized the importance of implementing targeted measures to reduce emissions from various sources, including transportation and industrial processes. To achieve this, it has recommended promoting cleaner transportation options, enforcing stricter emissions standards for point sources, and improving industrial efficiency. Another study [10] has highlighted the need for public awareness campaigns that encourage individuals to adopt behaviors that help reduce their contribution to air pollution. Overall, both studies underscore the importance of taking a multifaceted approach to address the sources of air pollution emissions and raise public awareness about the adverse impacts of air pollution on health and the environment in Kazakhstan.

The study [11] emphasizes the significance of emissions and meteorological conditions in causing high levels of air pollution in urban areas. Specifically, the research focuses on Astana, a sharply continental climate characterized by long, cold winters and moderately dry, hot summers. The city's average annual summer and winter temperatures are approximately 20°C and −15°C, respectively, and summer heatwaves and winter frosts can result in extreme temperatures. The authors have highlighted the crucial role of regional climatic conditions, including wind, temperature, and pressure, as well as the ventilation characteristics of the atmospheric space in the accumulation of pollutants in the air, particularly in the surface layer of the atmosphere. For instance, during periods of stagnation of anticyclones, pollutant concentrations can remain high for extended periods. The study underscores the importance of considering emissions and meteorological factors when addressing air pollution in urban areas such as Astana.

Our study aims to examine the effect of emissions on air quality in Astana, determine the reliability of air pollution measurement data, and quantify the contribution of emissions to air pollution by combining monitoring and modeling techniques. The article studied temporal air quality patterns in Astana based on air pollution monitoring data from the US Embassy's fixed stationery and Sergek's traffic air emission sensors over several years. As a result, the reliability of data sources for obtaining data on air quality is determined.

Methodology

Conducting natural experiments is very expensive and, in many cases, practically impossible. Therefore, it is necessary to use advanced methods to assess, predict and prevent anthropogenic changes that entail fatal consequences. In solving this issue, an undeniable advantage belongs to mathematical and computer modeling, which allows considering the orographic and climatic features of the region, choosing the optimal conditions for the operation of industrial facilities, and correctly and reasonably formulating recommendations for taking measures aimed at improving the environmental situation [12].

DOI: 10.37943/13QOJG5081

According to the Methodology for the Formation of Environmental Statistics Indicators approved by Order No. 223 of the Acting Chairman of the Statistics Committee of the Ministry of National Economy of the Republic of Kazakhstan dated December 25, 2015, in Kazakhstan for calculation are used particulate matter, sulfur dioxide, carbon monoxide, nitrogen dioxide, phenol, and formaldehyde. Depending on the harm caused to health, the air pollution index is classified as a safe, elevated, unsafe, and dangerous level. Additionally, hazard classes of harmful substances were provided. In the study, two data sources were used to analyze and compare the air quality data: the data from the US Embassy's fixed air monitoring station and the data from Sergek's sensor network from many parts of the city. The analysis included the assessment of the duration of exposure of the population to the effects of elevated concentrations.

The calculation is based on the approved maximum permissible norms of harmful substances and the level of harmful substances according to the Methodology collected by the sensors. The calculations are made according to the following formula:

$$\frac{x_1}{x_2} * y_1 \tag{1}$$

where:

– x_1 is the average annual concentration of the substance in the atmosphere,

– x_2 is the average daily concentration, calculated by the average value for one day,

– y_1 is the coefficient of the hazard class of the harmful substance.

The number of micrograms (mcg) of pollutants in cubic meters (m3) of air is used as a unit of measurement of absolute values of concentrations of pollutants (1) [13].

Another notable aspect of this study entails the computation of the average value derived from considerable latitude and longitude points. In order to juxtapose data obtained from a stationary station with that obtained from an interconnected array of sensors, it becomes imperative to harmonize the information acquired from a network of sensors situated near the fixed station, thereby consolidating it into a unified format. When considering an essential arithmetic mean between two locations, it is crucial to acknowledge the limitations of directly averaging longitude and latitude coordinates. Although this approach may yield satisfactory outcomes in regions with lower latitudes, its effectiveness deteriorates significantly as one approaches higher latitudes and ultimately breaks down near the poles. To address this issue, our method involves converting longitude, and latitude coordinates into three-dimensional Cartesian coordinates (x, y, z). A resultant Cartesian vector is obtained by computing the average of these Cartesian coordinates. Subsequently, the vector is reconverted back into longitude and latitude coordinates. It is worth noting that the normalization of the vector might not be necessary for this averaging process, thereby permitting a straightforward summation to derive the desired average:

$$\begin{gathered} latitute = rad2\deg * \operatorname{atan2}(\mathrm{z}, \sqrt{(x^2 + y^2)}) \\ longitute = rad2\deg * \operatorname{atan2}(-\mathrm{y},\ x) \end{gathered} \tag{2}$$

where:

– $atan2(a, b)$ is the arctangent of a divided by b. This function takes two arguments and returns the angle, in radians, between the positive x-axis and the point $(b,\ a)$ in the Cartesian coordinate system,

– x, y, z is the Variables representing the Cartesian coordinates of a point,

– $rad2deg$ is a constant representing the conversion factor from radians to degrees.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 13, MARCH 2023

In general, when studying the degree of technogenic pollution of the atmospheric air in the city of Astana, the following set of methods will be used: statistical processing of long-term climatic data where the data collected from monitoring posts for the state of atmospheric air; probabilistic-statistical - when establishing correlation dependencies; computer modeling of areas of distribution of pollutants from industrial sources and vehicles; graphic and construction - when combining model results and city schemes; comparative when establishing the proximity of the calculated characteristics with observational data. The data obtained from leading sources should be used in the listed activities.

Data Analysis

Data from the air quality sensors of the US Embassy and Sergek in Astana were collected to draw a conclusive analysis. The primary objective was to compare the PM 2.5 measurement results from both datasets by utilizing the relevant embassy data as the basis.

Before normalizing the data, a preliminary processing and visualization stage was conducted to ensure the accuracy and reliability of the data. The data processing step involved identifying and addressing missing or outlier data points in the US Embassy and Sergek datasets. The US Embassy and Sergek datasets were prepared as data frames for convenience. Both data frames from these datasets were prepared and normalized using Python, a popular tool for data analysis and processing (Figure 1). Python is a well-known and widely used language for data analysis and processing due to its user-friendly interface, powerful data processing libraries, and wide range of data processing and data visualization tools.

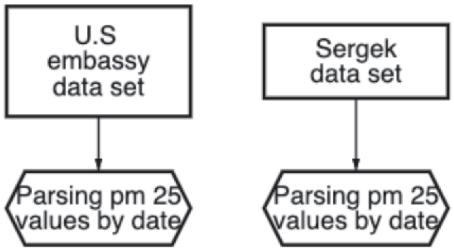

Figure 1. Selection of targeted properties from datasets

The US Embassy dataset is clean, ordered, and has no scatter, whereas the Sergek dataset is scattered and requires normalization. One of the peculiarities of the Sergek dataset is the recording of PM 2.5 levels emitted from passing cars, resulting in the collection of multiple air quality data at different times on a specific date.

Both datasets contain fields and values such as date and PM 2.5 emission value. To standardize the structure of the datasets and make them identical, normalization of the Sergek dataset was necessary. Normalization is a common data preprocessing technique to scale data to a standard range. This technique ensures that the data is comparable and can be analyzed more effectively. The input for the normalization process consisted of date-time format and various values of PM 2.5 for the day with a shift of 3-4 hours, while the output was the average PM 2.5 value for the entire day grouped by date. As a result of this process, PM 2.5 indicators were grouped by date, and the average value for the selected date range was displayed (Figure 2).

DOI: 10.37943/13QOJG5081

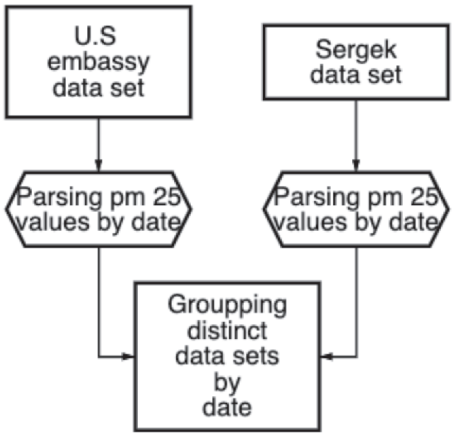

Figure 2. Data preparation

After normalizing data frames from the datasets, an analysis was conducted to compare the data obtained from each source. The first step of the analysis was to group the data by date, ignoring any spreads or differences in the time of measurements. Next, the difference in the measurement results between the sensors of each data frame was calculated and displayed. This allowed for a more detailed understanding of the differences between the two data sources and any potential discrepancies in the data. Overall, this analysis helped to identify any inconsistencies or outliers in the data and provided a more accurate picture of the air quality situation in the studied area (Figure 3).

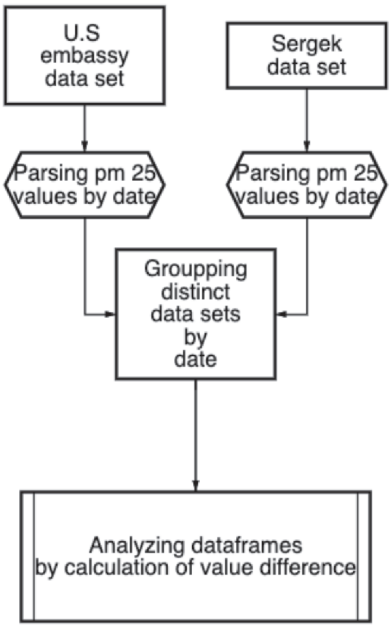

Figure 3. Data analysis

In the resulting action with the code, the results for the several month rates in the data frames are output using previously written methods (Figure 4).

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 13, MARCH 2023

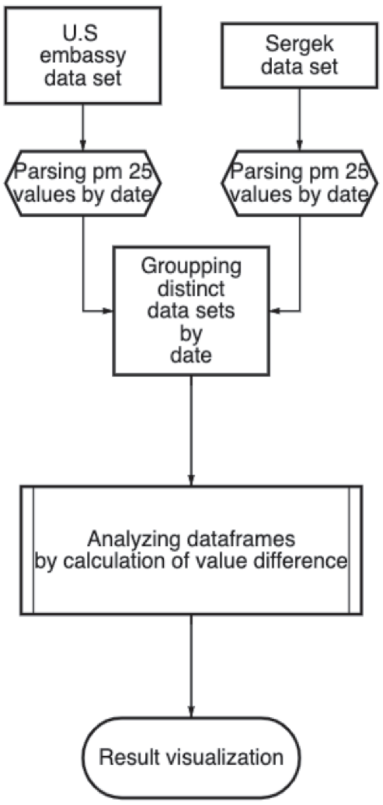

Figure 4. Algorithm Visualization

Let C be the concentration of a pollutant in the air, measured in units of parts per million (ppm). Let S be the reading obtained from the sensor or instrument measuring the pollutant concentration, measured in units of volts. Then it can be used a mathematical model to convert the sensor reading into the pollutant concentration as follows:

$$C = kS + b \tag{3}$$

where k is the calibration factor, which relates the sensor reading to the pollutant concentration, and b is the baseline value, representing any background pollution levels that may be present. The values of k and b can be determined through calibration experiments, where known concentrations of the pollutant are introduced into the air, and the corresponding sensor readings are recorded (3).

Real-time air quality monitoring can be achieved using sensors and other instruments that continuously measure pollutant concentrations. These measurements can be transmitted to a central database or dashboard, where they can be analyzed and visualized in real time. This allows decision-makers to quickly identify areas with high pollution levels and take appropriate action to protect public health and the environment.

In addition to real-time monitoring, long-term monitoring programs can provide valuable insights into trends and patterns in air quality over time. By collecting data over the years, we can identify changes in pollutant concentrations and track the effectiveness of pollution control measures. This information can inform policy decisions related to environmental protection and public health.

The formula for calculating the mean average for the given air quality data collected by the sensor throughout the day is:

DOI: 10.37943/13QOJG5081

$$mean\ average\ =\ (x1\ +\ x2\ +\ x3 + \ldots + xn)\ /\ n \quad (4)$$

where:

– $x1$ represents the value of the first data point,

– n represents the total number of data points.

– variables $x2$, $x3$, ..., xn represent the values of the subsequent data points.

The median provides a measure of central tendency representing the median value or the average of the two middle values in the sorted data set, which helps analyze skewed or non-normally distributed data (4).

In general, the visualization of the comparison between the average concentration levels of air pollutants and the indices of maximum permissible concentrations can provide valuable insights into the extent of air pollution in each area. It can help identify potential sources of pollution, assess the impact of existing pollution control measures, and inform the development of future policies and interventions to improve air quality (Table 1).

Table 1. Data representation with value difference of data sources.

Date	PM 25 U.S Embassy	PM 25 Sergek	Difference
11-06	71.0	27.8	43.2
11-07	63.0	33.9	29.1
11-08	61	20.9	40.1
11-10	87	32.1	54.9
11-11	121	41.3	79.7
11-12	155	36.3	118.7
11-13	64	38.6	25.4
11-14	63	10.5	52.5
11-15	60	10.5	49.5
11-16	62	10.5	51.5
11-17	63	10.5	52.5
11-18	100	10.5	89.5
11-19	74	10.5	63.5
11-20	71	10.5	60.5
11-21	68	10.5	57.5
11-22	64	10.5	53.5
11-23	59	10.5	48.5
11-24	66	10.5	55.5
11-25	61	10.5	50.5
11-26	62	10.5	51.5
11-27	70	10.5	59.5
11-28	60	10.5	49.5
11-29	62	7.9	54.1
11-30	76	15.1	60.9
12-06	71	25.5	45.5
12-07	68	48.3	39.7
12-08	59	19.6	39.4
12-10	60	188.1	119.9
12-11	77	23.5	53.5
12-13	113	40.3	72.7

Scientific Journal of Astana IT University
ISSN (P): 2707-9031 ISSN (E): 2707-904X
VOLUME 13, MARCH 2023

12-14	83	40.3	42.7
12-15	120	40.3	70.7
12-16	82	40.3	41.7
12-17	65	40.3	24.7
12-18	110	40.3	69.7
12-19	72	40.3	31.7
12-20	167	89.9	76.1
12-21	136	247.2	110.8
12-22	131	40.9	90.1
12-23	110	40.9	69.1
12-24	153	10.4	142.6
12-25	179	154.4	24.6
12-26	240	154.4	85.6
12-27	195	61.9	133.1
12-28	90	61.9	28.1
12-29	73	61.9	11.1
12-30	75	61.9	13.1

The results show that the PM2.5 data from Sergek sensors are very different from the data from the US Embassy. Therefore, the Sergek data should be perceived as something other than relevant. The reliability of the data studied using methods of mathematical statistics is 87-90%.

Climatic characteristics of the city during the year do not provide complete dispersion of anthropogenic emissions in the atmospheric air; the city is characterized by a high degree of pollution, and a significant excess of maximum permissible concentrations is associated with a high level of industrial development of the city and the proximity of the territorial location of enterprises to the residential sector. For the first time, the zoning of the city territory was carried out according to the duration of elevated concentrations in the atmospheric air based on model calculations, which indicates that the city's population is exposed to the adverse effects of harmful ingredients for a long time.

Data from the US Embassy indicate changes in the composition of the air depending on the time of year, day, and based on global climate change. It can be concluded that, in general, the level of pollution increases during the cold heating season, whereas in summer, the pollution level is much lower.

Sergek's data are calculated mathematically, which excludes the assumption of a mechanical accounting error, but at the same time, misses essential features such as vehicle characteristics, the climate of the region, and the time of year.

Acknowledgement

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. BR10965311 "Development of the intelligent information and telecommunication systems for municipal infrastructure: transport, environment, energy and data analytics in the concept of Smart City").

Conclusion

Modeling of pollution processes based on atmospheric monitoring allows for the most accurate solving such tasks as the placement of environmental monitoring posts, evaluating the contribution of individual industrial facilities to the pollution of residential areas, predicting adverse situations in the distribution of emissions, choosing a site for the construction of a

specific facility (school, residential building, stadium or industrial site), develop evacuation plans in the event of salvo emissions, as well as many other tasks. All this can be solved in real-time, and management decisions can be made promptly using mathematical modeling methods to study processes and phenomena. However, it is worth considering that this activity is possible only if reliable data on the air conditions of the studied region is available.

As the analysis showed, despite a large amount of information about air pollution collected in Kazakhstan, there is a severe discrepancy in the actual data in which one may encounter an «environmental error» - a formal error in the interpretation of statistical data, which can lead to unfortunate consequences.

References

1. Dehghani, S., Vali, M., Jafarian, A., Oskoei, V., Maleki, Z., & Hoseini, M. (2022). Ecological study of ambient air pollution exposure and mortality of cardiovascular diseases in elderly. *Scientific Reports*, *12*(1), 21295. https://doi.org/10.1038/s41598-022-24653-0
2. Akhatova, A., Kassymov, A., Kazmaganbetova, M., & Rojas-Solórzano, L. (2015). CFD simulation of the dispersion of exhaust gasses in a traffic-loaded street of Astana, Kazakhstan. *Journal of Urban and Environmental Engineering*, *9*(2), 158-166. https://doi.org/10.4090/juee.2015.v9n2.158166
3. WiseVoter. (n.d.). *Most Polluted Countries in the World*. Retrieved January 31, 2023, from https://wisevoter.com/country-rankings/most-polluted-countries
4. Elorda. (2022, May 31). *Zagryazneniye vozdukha v Kazakhstane – prichina bolee 10 tysyach prezhdevremennykh smertey i yezhegodnogo ushcherba na $10,5 milliarda* [Air pollution in Kazakhstan is the cause of more than 10,000 premature deaths and an annual damage of $10.5 billion]. Retrieved January 31, 2023, from https://elorda.info/sotsium/18422-1654002356/
5. Nazarenko, Y., Pal, D., & Ariya, P. A. (2021). Air quality standards for the concentration of particulate matter 2.5, global descriptive analysis. *Bulletin of the World Health Organization*, *99*(2), 125–137D. https://doi.org/10.2471/BLT.19.245704
6. Kenessary, D., Kenessary, A., Adilgireiuly, Z., Akzholova, N., Erzhanova, A., Dosmukhametov, A., Syzdykov, D., Masoud, A. R., & Saliev, T. (2019). Air Pollution in Kazakhstan and Its Health Risk Assessment. *Annals of global health, 85*(1), 133. https://doi.org/10.5334/aogh.2535
7. Informburo.kz. (2021, October 19). *Zhiteli stolicy zadykhayutsya ot smoga: Gorod tak i ne gazificirovali* [Residents of the capital are suffocating from smog: The city was never gasified]. Retrieved from https://informburo.kz/stati/zhiteli-stolicy-zadyhayutsya-ot-smoga-gorod-tak-i-ne-gazificirovali
8. Helmut, Mayer. (1999). Air pollution in cities. *Atmospheric Environment*, *33*(24), 4029-4037. https://doi.org/10.1016/S1352-2310(99)00144-2
9. Munir, S., Mayfield, M., Coca, D., Mihaylova, L. S., & Osammor, O. (2020). Analysis of Air Pollution in Urban Areas with Airviro Dispersion Model—A Case Study in the City of Sheffield, United Kingdom. *Atmosphere*, 11(3), 285. https://doi.org/10.3390/atmos11030285
10. Yerzhanova, A., Aitkhozhina, N., Abuduwaili, J., & Yermukhanova, A. (2021). Industrial Emissions Trends and Air Pollution Levels in Industrial Cities in Kazakhstan. *Atmosphere*, *12*(3), 314. https://doi.org/10.3390/atmos12030314
11. Ormanova, G., Karaca, F., & Kononova, N. (2020). Analysis of the impacts of atmospheric circulation patterns on the regional air quality over the geographical center of the Eurasian continent. *Atmospheric Research*, 237, 104858. https://doi.org/10.1016/j.atmosres.2020.104858
12. Khashirova, T.Y., Akbasheva, G.A., Shakova, O.A., & Akbasheva, E.A. (2017). Modelirovanie zagryazneniya atmosfernogo vozdukha [Modeling of air pollution]. *Fundamental'nye Issledovaniya* [Fundamental Studies], 2017, 8-2. Retrieved from https://fundamental-research.ru/ru/article/view?id=41669
13. The Law of the Republic of Kazakhstan No. 174-II «*On Currency Regulation and Currency Control*», adopted on December 17, 2003. Retrieved from https://adilet.zan.kz/rus/docs/V1500012931