Statistical Analysis of Measuring Errors the Pollution of the Atmospheric Bottom Layer by Exhaust Gas

Iryna Darmorost[1], Petro Stakhiv[2], Mykola Shynkaryk[3], Vasylyna Melnyk[4]

1,3,4. Department of Computer Science, Ternopil National Economic University, UKRAINE, Ternopil, 8 Chekhova str., email: imadiudia@gmail.com[1]

2. Lodz University of Technology, POLAND, Lodz, 116 Stefana Żeromskiego str., email: spg@polynet.lviv.ua

Abstract: **The statistical analysis of the errors of measurements pollution in the bottom-layer of the atmosphere with exhaust gases of vehicles has been carried out. The method of performing statistical analysis is proposed. On this basis, the most optimal time averaging interval for the instantaneous concentrations of harmful substances is established.**

Keywords: **traffic intensity, harmful emissions, nitrogen dioxide, correlation coefficient, random error, averaging interval.**

I. Introduction

The progress of human society is inseparable from the history of transport development. With the expansion of states, the construction of cities, with the resettlement of people in increasingly large territories, the growth of trade rates of development of transport has steadily increased. The rapid pace of growth in the level motorization of the population leads to inevitable negative changes in the environment. Pollution of the atmosphere by harmful emissions of vehicles causes irreparable harm to the health of the population. It is known that during the day, people consume about 15-25 kg of air, 2.5-5 kg of water, 2.5 kg of food. When inhaled, the chemical elements are absorbed by the body most intensively. Thus, lead, which comes with air, is absorbed by blood by 60%, if it comes with water, it is absorbed by 10%, with food - by 5%. Therefore, when polluting the environment, atmospheric air is the main supplier of toxic substances in the human body [1].

The exhaust gases of cars contain more than 200 compounds and components, many of which are very toxic.

The environment contains carbon monoxide (CO), nitrogen (NO), sulfur dioxide (SO2), aldehydes, soot (C), lead (Pb), and others [2]. One of the most harmful substances, the concentration of which exceeds any allowable norms is nitrogen dioxide, the chemical formula of which – NO_2.

There is a clear linear dependence between the traffic intensity and the concentration of harmful substances in the atmosphere. However, measuring the value of NO_2, with the use of modern means of measurement, occurs almost instantaneously, and in this connection, the problem of averaging the obtained indicators. In the paper [3], experimental studies were performed and the correlation coefficients were calculated for different time averaging intervals: for 5, 10, 20 and 30 minutes. The correlation coefficient is the most optimal at the averaging interval in 20 minutes. In this work it is proposed to investigate the statistical characteristics of random error, in order to confirm the effectiveness of the selected averaging interval.

II. Statement of the Problem

The traffic intensity is one of the most important factors contributing to the pollution of the bottom-layer of the atmosphere by harmful emissions, but the concentration of air in a chemical compound such as nitrogen dioxide varies depending on the characteristics of the medium in which the measurements are carried out.

Based on preliminary studies, it was found that the traffic intensity and the known concentrations of harmful substances in the atmosphere in certain time bands correlate. The research was based on the assumption that the random error is normally distributed with zero mathematical expectation, and as a result of averaging, is compensated for this error.

And on this basis the idea was based on the choice of averaging interval.

It has been established that in order to compensate for the casual component related to the influence of other factors on pollution of the city's territory, such as ventilation, it is advisable to choose the range of averaging instantaneous values of the concentration harmful emissions and the traffic intensity of the transport units close to the interval of 20 minutes. Such an interval is used in standard measuring techniques Sanitary and Epidemiological Services.

In order to confirm or refute the justified averaging interval, it is proposed to investigate the statistical characteristics of the random error.

As is known, the correlation coefficient between the two variables is given by Eq.1:

$$r_{xy} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{m}(x_i - \bar{x})^2 \sum_{i=1}^{m}(y_i - \bar{y})^2}}, \quad (1)$$

where $\bar{x}, \bar{y}$ – average sample from each sample $\vec{x} = x_1 ... x_m ; \vec{y} = y_1 ... y_m$ accordingly, $r_{xy} \in [-1,1]$.

It is worth mentioning that the values $\vec{x} = x_1 ... x_m ; \vec{y} = y_1 ... y_m$, are calculated for different time averaging intervals of instantaneous values of traffic intensity and concentration of harmful substances, respectively.

Thus, for different time averaging intervals of instantaneous values, we calculate the correlation coefficient between the traffic intensity and the concentration of nitrogen dioxide and carry out a statistical analysis of the errors measurements of contamination the bottom-layer of the atmosphere by exhaust gases of vehicles.

II. Method of Researching

Statistical analysis of the measurement errors the bottom-layer of atmosphere on the example of the established correlation between the traffic intensity and the concentration of nitrogen dioxide (NO_2) was conducted.

To obtain experimental data, measurements of the concentration of nitrogen dioxide in the air at the crossroads of streets in city Ternopil Chekhova - Za Rudkoiu was conducted. Average air temperature was 0 ° C, humidity - 70%. Obtaining a sample of data was done using gas analyzer SPEC Sensors, DSG – NO2- 968-037, accuracy - ±15%, at a range of operating temperatures from -20 ºC to +40 ºC. The traffic intensity was evaluated using a set webcam every minute. Concentration value NO_2 received every second.

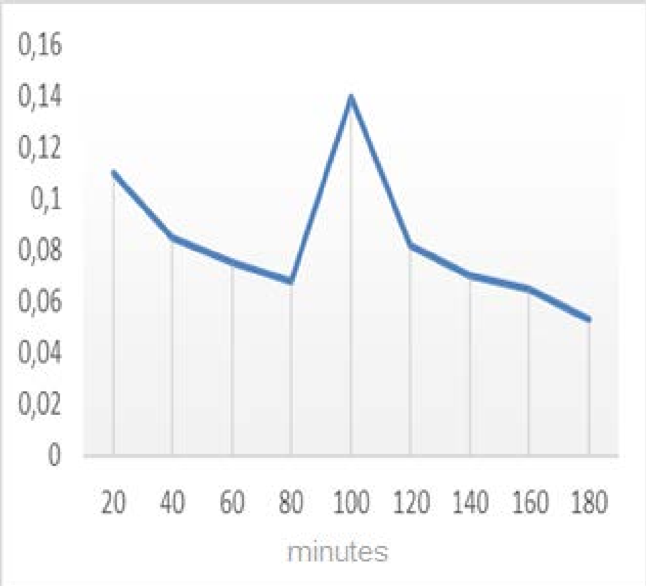

Fig.1. Interval of averaging concentration of NO_2 in 20 minutes.

Using eq. (1), the correlation coefficients were calculated on time averaging intervals of 5, 10, 20, and 30 minutes. Figure 1 shows a graph of the averaging interval of nitrogen dioxide concentration in 20 minutes [3].

Figure 2 shows the generalized results of the correlation coefficients at different time intervals.

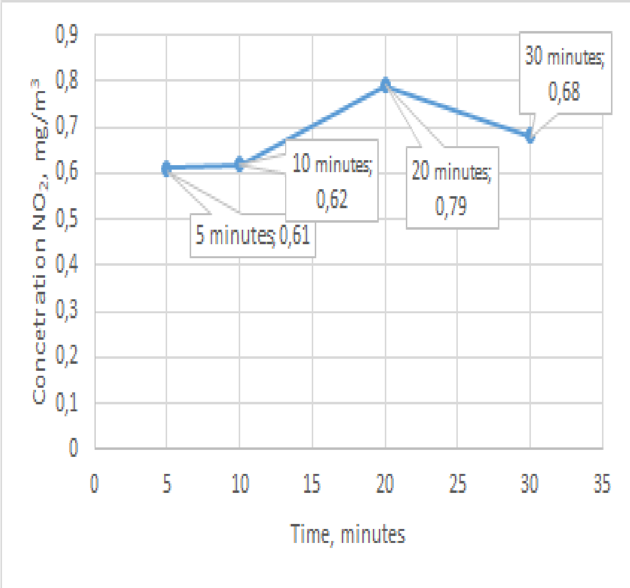

Fig.2. Coefficients of correlation for different time averaging intervals.

As you can see from Figure 2, we get a curve that has an extremum at the point with averaging of 20 minutes. For statistical analysis, the averaging intervals of 1, 5, 10, 20 and 30 minutes were selected.

The method of conducting the analysis includes the following steps:

1. Averaging of the concentrations of nitrogen dioxide in the bottom-layer of atmosphere at intervals of 1, 5, 10, 20, 30 minutes, in a time interval of 1 hour. For each interval averaging according to the value of the interval. For each of the intervals, 60, 12, 6, 3, and 2 averaging values were obtained, respectively.

2. On each averaging interval, net error and average value were selected. The calculation of the values of a random error occurs according to the following equations:

$$e_i = a_{mid} - z_m, \tag{2}$$

where e_i - random errors at different time intervals, a_{mid} - averaging the concentration of nitrogen dioxide in different time ranges, z_m - value measurement of nitrogen dioxide per second, $m=1..n$.

3. At all averaging intervals after calculations using eq. (2), the sum of random errors was obtained by the equations:

$$S_i = \sum_{i=1}^{N} e_i \ , \tag{3}$$

where S_i - amount of error values at different time averaging intervals.

III. Analysis of Received Results

The results of the research are shown in Figures 3-7. Statistical analysis of measurement errors, according to the research method, was conducted for different time intervals, in particular: 1, 5, 10, 20 and 30 minutes.

Figure 3 shows the value of random errors in the averaging interval of 1 minute.

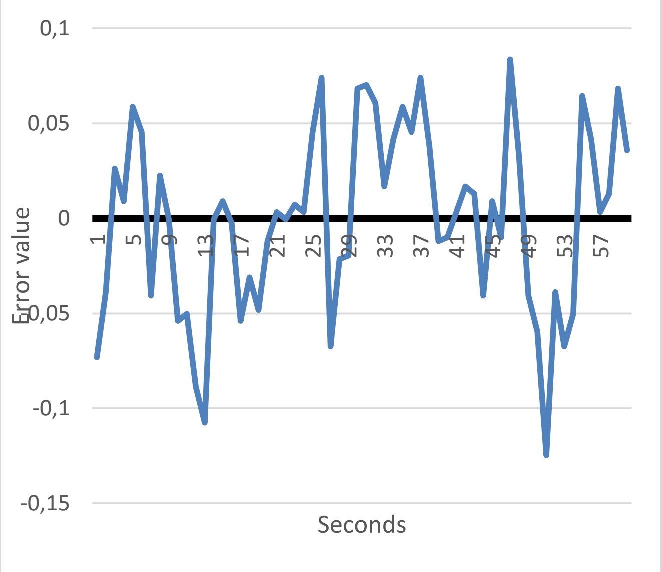

Fig. 3. The value of random errors in the time range of 1 minute in the averaging interval in 1 minute.

Figure 4 shows the value of random errors in the averaging interval in 5 minutes.

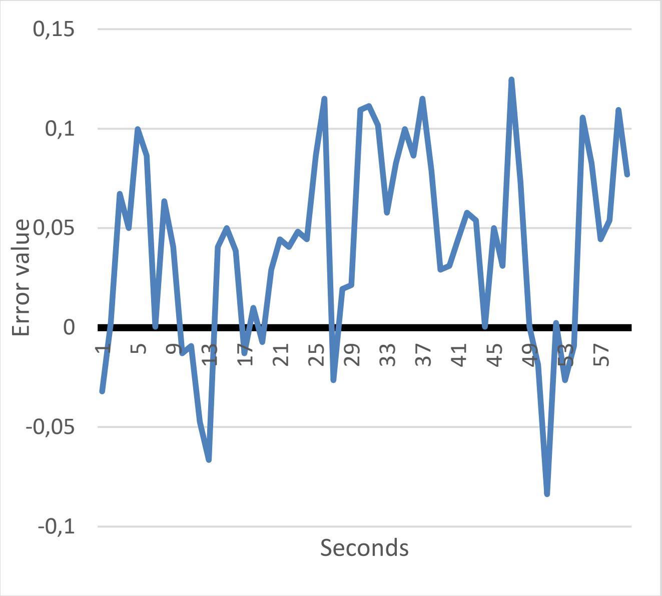

Fig. 4. The value of random errors in the time range of 1 minute in the averaging interval of 5 minutes.

Figure 5 shows the values of random errors in the averaging interval of 10 minutes.

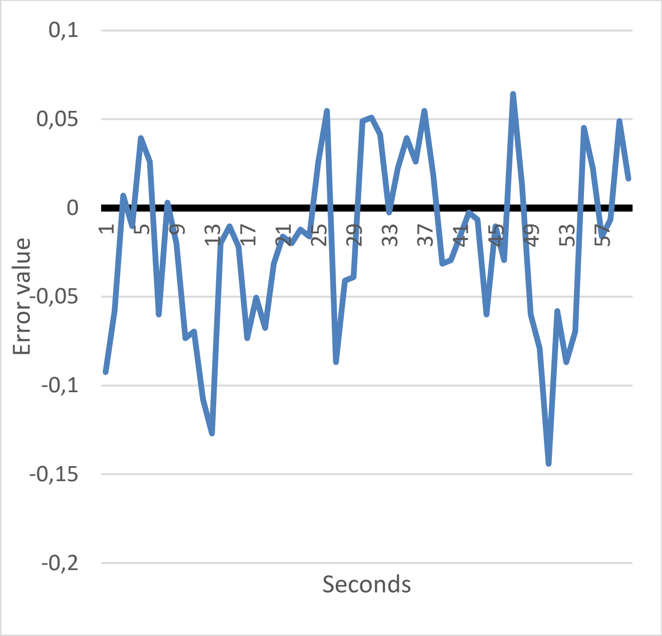

Fig. 5. The value of random errors in the time range of 1 minute in the averaging interval in 10 minutes.

Figure 6 shows the value of random errors in the averaging interval in 20 minutes.

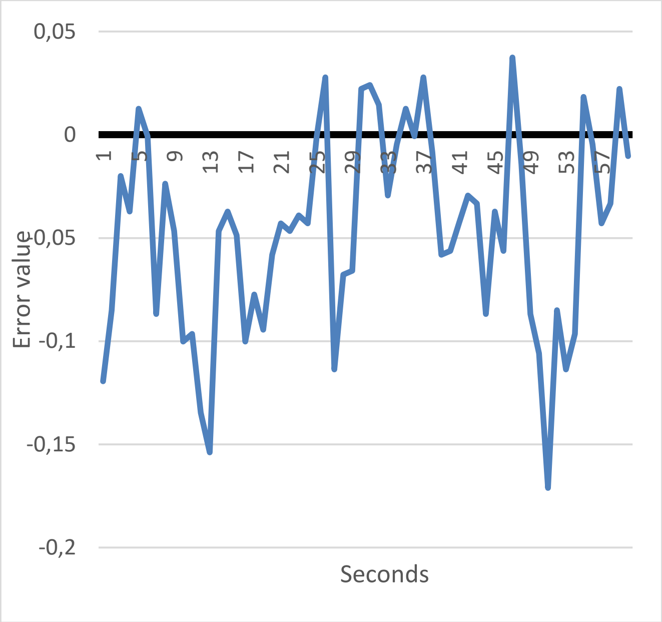

Fig. 6. The value of random errors in the time range of 1 minute in the averaging interval in 20 minutes.

Figure 7 shows the value of random errors in the averaging interval in 30 minutes.

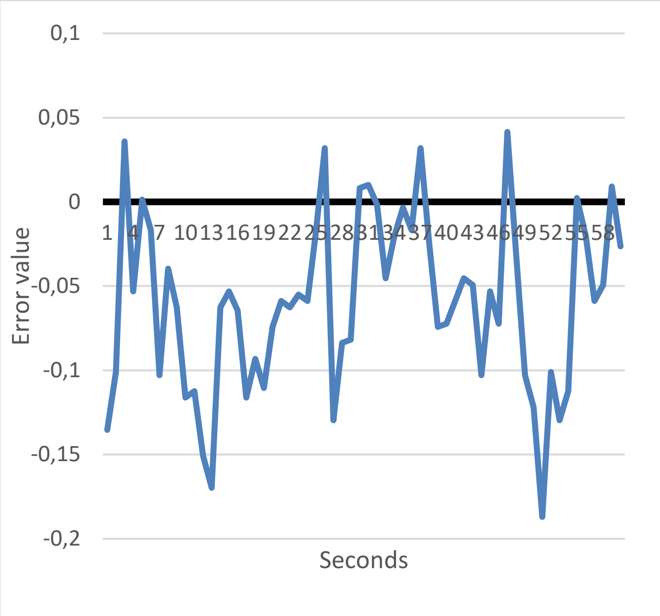

Fig. 7. The value of random errors in the time range of 1 minute in the averaging interval in 30 minutes.

On the basis of the proposed error, using the eq. (3) the general errors are calculated.

On the interval in 1 minute $S_1 = -7{,}015$, it means that the interval error is not normally distributed, but includes trend values, on 5 minutes $S_5 = 13{,}148$, which is also not the optimum value, in 10 minutes - $S_{10} = 2{,}94$, on 20 minutes - $S_{20} = 2{,}808$, on 30 minutes - $S_5 = 3{,}0056$.

Apparently, as a result of the calculations performed by the proposed methodology, it has been found that in subsequent studies, as averaging interval of , it is necessary to use the averaging interval in the range of 10 to 20 minutes.

The vast majority of sources [4] use an interval of averaging of 1 hour. The averaging graph for 1 hour is depicted in Figure 8. For the visualization, a sample of data was obtained from the measurement of nitrogen dioxide concentration values at Street Jana Pawla II in Lodz, Poland. On the ordinates axis, the averaged values are deferred value of nitrogen dioxide (NO_2), along the abscissa axis - hours of day.

The conducted studies have shown that such averaging gives distortion of results due to the use of an incorrect and ineffective averaging interval of the measured values of nitrogen dioxide, which can not be neglected by the action of random factors such as ventilation of the environment, that is, gusts of wind, vertical and horizontal streams, turbulence, and so on.

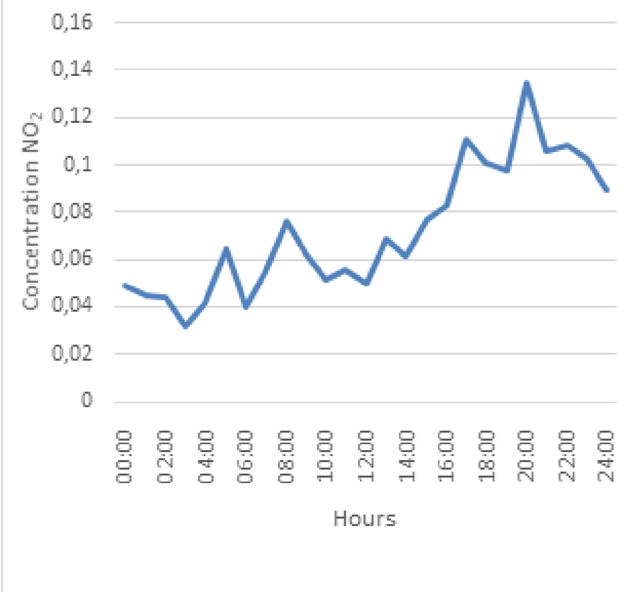

Fig. 8. Measured values of nitrogen dioxide in the averaging interval in 1 hour.

IV. Acknowledgment

This research was supported by National Grant of Ministry of Education and Science of Ukraine “Mathematical tools and software for control air pollution from vehicles” (0116U005507).

V. Conclusion

A statistical analysis of random errors was carried out at different intervals of averaging - at 1, 5, 10, 20 and 30 minutes. Thus, the averaging interval of instantaneous values of the concentration of nitrogen dioxide in bottom-layer of atmosphere was formed, which will be used in further studies to construct models of the dependence the concentration of harmful emissions between the traffic intensity . The value of the averaging interval will be in the range of 10 to 20 minutes.

References

[1] V. Stukanov, “Effect of motor transport on the state of the environment of the large industrial cities”, *Vestnik VSU, Series: Chemistry, Biology, Pharmacy*, no.1,pp.168-175, 2012 (In Ukrainian).

[2] V.Garin, “Industrial ecology”, *Moscow-Marshrut*,328 p., 2005 (In Russian).

[3] M.Dyvak, I.Darmorost, R.Shevchuk, V.Manzhula and N.Kasatkina., “Correlation analysis traffic intensity of the motor vehicles and the air pollution by their harmful emissions”, *Proceedings of XIVth International Conference on Modern Problem of Radio Engineering, Telecommunications and Computer Science (TCSET)*, Lviv-Slavske, 2018, pp.855-858.

[4] Wojewódzki Inspektorat Ochrony Środowiska w Łodzi [Online]. Available: http://www.wios.lodz.pl/. Accessed on: April, 22, 2018.